Doctoral theses at NTNU, 2022:148

Lin Jiang

Risk factors other than smoking for incidence of lung cancer: The influence of a sedentary lifestyle, asthma, and body mass index

Prospective cohort and Mendelian randomization studies based on data from The HUNT Study

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of Philosophiae Doctor
Faculty of Medicine and Health Sciences
Department of Public Health and Nursing

NTNU
Norwegian University of Science and Technology

Lin Jiang

Risk factors other than smoking for incidence of lung cancer: The influence of a sedentary lifestyle, asthma, and body mass index

Prospective cohort and Mendelian randomization studies based on data from The HUNT Study

Thesis for the Degree of Philosophiae Doctor

Trondheim, June 2022

Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Public Health and Nursing

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Medicine and Health Sciences
Department of Public Health and Nursing

ISBN 978-82-326-5898-5 (printed ver.)
ISBN 978-82-326-5415-4 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2022:148

Printed by NTNU Grafisk senter

Risikofaktorer enn røyking og insidens av lungekreft: påvirkning av inaktiv livstil, astma og kroppsmasseindeks

Prospektive kohort og Mendelsk randomisering analyser av HUNT data

Lungekreft er en av de vanligste kreftformene med lave overlevelse over hele verden. Selv om røyking er den viktigste risikofaktoren for lungekreft, ses den vanligste typen, adenocarcinom, hos mange som aldri har røykt. Ettersom den største økning i insidens av lungekreft skyldes adenocarcinom, er det mulig at andre potensielle risikofaktorer enn røyking har en viktig rolle i utvikling av lungekreft.

Denne avhandlingen består av tre delstudier med overordnet mål om å avdekke mulige sammenhenger mellom inaktiv livstil, astma, kroppsmasseindeks (KMI) og insidens av lungekreft generelt og undergrupper. Datakilder er Helseundersøkelsen i Trøndelag (HUNT) og Kreftregisteret. De to første studiene er prospektive kohort-studier, mens den siste studien fokuserer på potensiell årsakssammenheng mellom KMI og forekomsten av lungekreft og undergrupper ved bruk av både observasjonelle- og Mendelsk randomisering (MR).

Den første studien viste ingen klar sammenheng mellom langvarig stillesitting og forekomsten av lungekreft, og småcellet (SCLC) eller ikke-småcellet (NSCLC) lungekreft. Men kombinasjonen av langvarig stillesitting og fysisk inaktivitet ser ut til å kunne øke forekomsten av lungekreft generelt og SCLC.

I den andre studien fant vi ingen klar sammenheng mellom astma generelt og forekomsten av lungekreft generelt eller undergrupper. Men delvis kontrollert astma så ut til å være assosiert med økt forekomst av lungekreft generelt og NSCLC. Blant voksne med aktiv astma fant man en tendens til økt forekomst av lungekreft generelt. Men det var ingen klar sammenhenger mellom ikke-aktiv astma eller kontrollert astma med forekomsten av lungekreft generelt og undergrupper.

I den observasjonelle delen av tredje studie, så KMI ut til å være omvendt assosiert med forekomsten av adenocarcinom i lungene. Men det ble ikke påvist dose responseffekt av 10 års endering i KMI og denne kreftformen. Dette tyder dermed på at det ikke er en årsakssammenheng. Ved multivariabel MR-analyse fant vi en positiv sammenheng mellom genetisk bestemt KMI og forekomsten av adenocarcinom når vi justerte for genetisk bestemt røyking. Men vi fant ingen klar sammenheng mellom KMI og forekomsten av øvrige undergrupper av lungekreft, enten i observasjonelle eller MR-analyser.

Avhandlingen viser at personer med mest inaktiv livstil har økt risiko for lungekreft. Tilsvarende økt risiko fant man også for personer med delvis kontrollert astma. Imidlertid kan vi ikke utelukke rest-konfundering av røyking i de to første studiene. Motstridende resultat mellom observasjonelle og MR-analyser tyder på at den påviste inverse sammenheng mellom KMI og adenocarcinom ikke er kausal. Det er nødvendig med flere MR studier for å avdekke forklaring for denne assosiasjonen i fremtiden.

Navn kandidat: Lin Jiang

Institutt: Institutt for samfunnsmedisin og sykepleie

Hoved veileder: Professor Xiao-Mei Mai

Biveileder: Senior forsker Yi-Qian Sun, Professor Arnulf Langhammer, Førsteamanuensis Ben Michael Brumpton, Professor Yue Chen

Finansieringskilde: Stiftelsen Dam gjennom kreftforeningen, Samarbeidsorganet mellom Helse Midt-Norge og NTNU

Ovennevnte avhandling er funnet verdig til å forsvares offentlig

for graden Ph.d. i medisin

Disputas finner sted i auditorium MTA i Fred Kavli-bygget

Fredag 10 juni 2022, kl. 12:15

Contents

Summary

Lung cancer is one of the most common cancer types with a low survival rate worldwide. Although smoking is the most important risk factor for lung cancer, many people with lung adenocarcinoma, the most common lung cancer subtype, are non-smokers. Given the increasing incidence of lung cancer is mainly accounted by adenocarcinoma, other potential risk factors than smoking may play important roles for the development of lung cancer.

The current thesis consists of three studies with the overall aim of exploring possible associations between a sedentary lifestyle, asthma, body mass index (BMI) and the incidence of lung cancer overall and its subtypes, using data from The Trøndelag Health Study (HUNT) and the Cancer Registry of Norway. The first two studies are prospective cohort studies, while the last one focuses on the potential causal association between BMI and lung cancer incidence using both observational and Mendelian randomization (MR) methods.

In the first study, we found that prolonged sitting was not independently associated with incidence of lung cancer overall, small cell lung cancer (SCLC) or non-small cell lung cancer (NSCLC). However, the combination of prolonged sitting and physical inactivity might increase the incidence of lung cancer overall and SCLC.

Although we did not observe a clear association between asthma overall and incidence of lung cancer overall or its subtypes in the second study, partially controlled asthma was suggested to be associated with an increased incidence of lung cancer overall and NSCLC. Adults with active asthma showed a tendency toward increased incidence of lung cancer overall. There was no clear association between non-active asthma or controlled asthma and incidence of lung cancer overall or its subtypes.

In the last study, BMI seemed to be inversely associated with the incidence of lung adenocarcinoma based on observational data. However, the BMI change results did not

support a dose-response relationship. By using a multivariable MR method, we further found a positive association between genetically determined BMI and the incidence of adenocarcinoma after genetically controlling for smoking. No clear associations were observed in other lung cancer subtypes either in observational or MR studies.

In summary, our results suggested that people who were mostly sedentary were at an increased risk of developing lung cancer. People who had partially controlled asthma also had an increased risk of lung cancer. However, residual confounding by smoking cannot be excluded completely in the first two studies. The discrepancy of findings between the observational and MR analysis suggested that the observed inverse association between BMI and lung adenocarcinoma might not be causal. More MR studies are needed to explore the nature of this association in the future.

Acknowledgements

My PhD project was carried out at the Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU) with the funding from the Stiftelsen Dam through the Norwegian Cancer Society (project no. 47090327) and top-up funding from the collaboration partner between the Liaison Committee for Education, Research and Innovation in Central Norway and Central Norway Regional Health Authority.

First of all, I will thank my main supervisor, Prof. Xiao-Mei Mai for opening the door of Epidemiology to me and giving me opportunity to learn a lot from her. Without her encouragement and support, I might give up my dream during my hardest time of life before I started my PhD. I am so grateful to her outstanding supervision. She is always available and patient, even during the lock-down periods due to coronavirus; She is always willing to share her expertise and encourage me to challenge myself. She is not only my supervisor, but also my mentor during the life. Her valuable advice helps me to improve my ability to do research and make me to be a better person. I cannot express how thankful I am for her with words.

Then, I wish to express my sincere gratitude to my co-supervisors and co-authors. I thank Senior researcher Yi-Qian Sun for his outstanding assistance to all my statistical challenges and his wonderful do-files and constructive suggestions when I am doing analyses for Mendelian randomization using R. I want to thank Professor Arnulf Langhammer for his valuable suggestions on how to use HUNT data in a proper way and feedback on all of my research papers. I am grateful to Associate Professor Ben Michael Brumpton for coming up new ideas to perform analyses and valuable academic discussions. My thank goes to Professor Yue Chen and Tom IL Nilsen for their great epidemiological knowledge and scientific advice. I want to thank researcher Linda Leivseth for her constructive suggestions for scientific writing and sharing her expertise in respiratory epidemiology. I also wish to express my gratitude to Dr. Sissel Gyrid Freim Wahl for her great suggestions in clinic aspects.

I want to acknowledge the HUNT Research Centre and all the participants that have contributed to the data I used in this thesis. I would also like to thank the Norwegian Cancer Registry for providing the crucial data on lung cancer.

Most importantly, I would thank my husband, Jie Wu for his encouragement and support all the time. Without him, I cannot make my dream come true. I will also thank my parents and parents-in-law for their indispensable support to the family during my busiest time of PhD study.

This thesis is dedicated to my wonderful sons, Yuhan and Erik. Their smiles are always my motivation to do better.

List of Figures

List of Tables

List of abbreviations

BMI: Body mass index

CI: Confidence interval

COPD: Chronic obstructive pulmonary disease

DAG: Directed acyclic graph

FEV1: Forced expiratory volume in first second

FVC: Forced vital capacity

GIANT: Genetic Investigation of Anthropometric Traits consortium

GINA: Global Initiative for Asthma

GOLD: Global Initiative for Chronic Obstructive Lung Disease

GRS: Genetic risk score

HR: Hazard ratio

HUNT: The Trøndelag Health Study

ICD-10: International Classification of Disease,10^{th} Edition

ICD-O: International Classification of Diseases of Oncology

ICS: Inhaled corticosteroids

IR: Incidence rate

IVW: Inverse variance weighted method

METs: Metabolic equivalences

MR: Mendelian Randomization

MR-PRESSO: MR- Pleiotropy Residual Sum and Outlier

NSCLC: Non-small cell lung cancer

RCT: Randomized controlled trials

PA: Physical activity

SCLC: Small cell lung cancer

SNPs: Single nucleotide polymorphisms

WHO: World Health Organization

List of papers

Paper I: Jiang L, Sun YQ, Brumpton BM, Langhammer A, Chen Y, Nilsen TIL, Mai XM. Prolonged sitting, its combination with physical inactivity and incidence of lung cancer: prospective data from the HUNT Study. *Front Oncol. 2019; 9:101*

Paper II: Jiang L, Sun YQ, Brumpton BM, Langhammer A, Chen Y, Nilsen TIL, Leivseth L, Wahl SGF, Mai XM. Asthma and asthma symptom control in relation to incidence of lung cancer in the HUNT study. *Sci Rep. 2021;11(1):4539*

Paper III: Jiang L, Sun YQ, Brumpton BM, Langhammer A, Chen Y, Mai XM. Body mass index and incidence of lung cancer in The HUNT Study: Using observational and Mendelian randomization approaches. *BMC Cancer. In review*

1 INTRODUCTION

1.1 Lung cancer and histologic types

Lung cancer is a disease in which abnormal cells uncontrollably grow in the tissues of the lung. Small cell lung cancer (SCLC) and non-small cell lung cancer (NCSLC) are two major subtypes of lung cancer. SCLC is an aggressive (fast-growing) cancer that forms at the center of the lung and can spread to other parts of the body [1]. The cancer cells are small and oval-shaped [1]. NCSLC is more common and can be categorized into adenocarcinoma, squamous cell carcinoma, and large cell carcinoma based on how the cell appearance [2]. Adenocarcinoma is a malignant tumor of glandular cells and is seen in breast, prostate and gastrointestinal tract, and has become the predominant subtype of lung cancer, accounting for 35–40% of all lung cancer cases [3]. More women than men have adenocarcinoma (42% vs 28%), while opposite pattern is found for SCLC (25% of women vs 44% of men) [3].

1.2 Epidemiology of lung cancer

From a worldwide perspective, lung cancer has been one of the most common cancer types for several decades. In 2020, there were an estimated 2.2 million new cases, which represented 11.4% of all new cases of cancer [4]. Lung cancer is also the most lethal cancer type in the world. In 2020, about 1.8 million people died from lung cancer, making it responsible for 18% of global deaths [4]. Women seemed to be less likely to die from lung cancer than men, with an estimate of around 0.6 million deaths in 2020 [4]. However, both incidence and mortality among women continue to rise in most western countries [5].

By the time they are diagnosed, more than half of people have already progressed to an advanced stage with metastasis, leaving them with a lack of effective treatment options [6]. Thus, the five-year survival rate of lung cancer was earlier reported as just 13% for people who were diagnosed between 2000–2007 in Europe [7]. Due to improvement in treatment, mainly for NSCLC, the latest after-diagnosis 5-year survival rate for lung cancer has been

increased [5]. In the US, the after-diagnosis 5-year survival rate was reported as around 20.5% between 2010–2016 [5]. Recently, mortality rate of NSCLC has been found to decline faster than its incidence in the US [8].

In Norway, lung cancer is the second most common cancer for both men and women after prostate and breast cancer [9]. As in other western countries, there is a gender difference in the incidence and mortality rate of lung cancer in Norway (Figure 1). For Norwegian men, the incidence and mortality rate of lung cancer stabilized in the mid-1990s before declining, but Norwegian women have continued to be diagnosed with, and die from lung cancer at an increasing rate over the past two decades [9]. Although the after-diagnosis five-year survival rate of lung cancer is low, it has increased by around 10% over the last 10 years in Norway [9], with women having a better survival rate than men (29% vs 23%) [9].

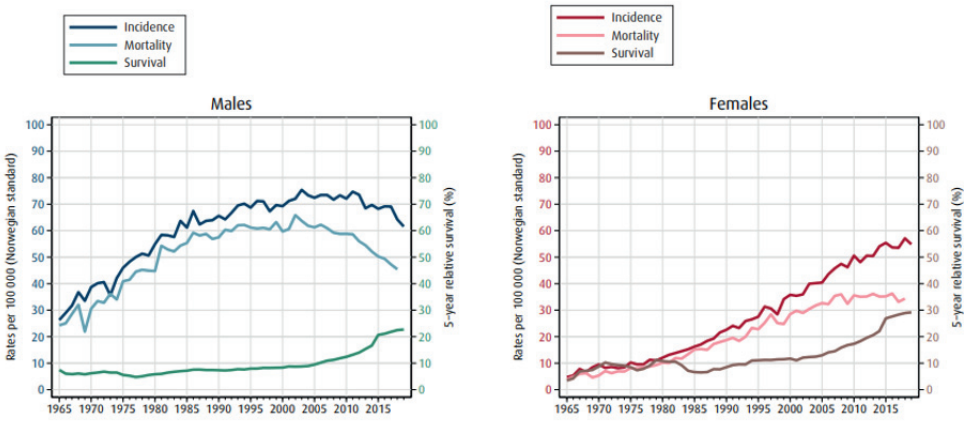


Figure 1. Trends in incidence, mortality rates and five-year relative survival rates in Norway since 1965, adapted from the Cancer Registry of Norway's 2019 Report on Cancer [9].

1.3 Smoking and lung cancer

Smoking is the most important risk factor for developing lung cancer. It is reported that over 80% of all lung cancer cases are due to smoking [10]. The influence of smoking on the development of lung cancer depends on smoking burden (the number of cigarettes smoked

daily, pack-years of smoking), and time since smoking cessation [7]. Although women seemed to have similar risks of lung cancer compared with men, it suggested that the risk of lung cancer due to smoking in women might be underestimated in a meta-analysis of 99 cohort studies [11]. In the meta-analysis, prevalence and intensity of smoking were higher in men compared with women in most included studies. Thus women might have a higher risk of lung cancer when the smoking epidemic reached full maturity in women [11]. In addition, there is a gender difference in susceptibility to smoking with a higher susceptibility in women [5, 7]. Compared with male heavy smokers, women with comparable smoking histories seem to have a higher risk of developing lung cancer [12]. With regard to histology, smoking is strongly associated with an increased risk of SCLC, while its association with NSCLC (especially adenocarcinoma) is weaker [7]. During the last decades, the occurrence of SCLC has decreased in western countries, probably due to successful tobacco legislation [7]. However, the proportion of adenocarcinoma is rising, with many people who are diagnosed reporting that they have never smoked (10–40%) [13]. This suggests that risk factors other than smoking may play an important role in the development of lung cancer, especially for NSCLC.

1.4 Risk factors other than smoking for lung cancer

Apart from smoking, passive smoking, domestic biomass fuels, chronic obstructive pulmonary disease (COPD) and other pulmonary conditions, ambient air pollution, and other environmental exposures, such as radon, asbestos, and workplace carcinogens, are also well-known risk factors for lung cancer [7]. Below, some potential risk factors that are not well studied are considered.

1.4.1 Sedentary lifestyle and leisure-time physical activity

The term “sedentary lifestyle” describes a series of human behaviors requiring low energy expenditure (≤1.5 metabolic equivalents [METs]) in a sitting or reclining posture when

awake [14]. This kind of lifestyle is highly prevalent in Western countries [15, 16]. More than one third of the whole population aged 15 years and over in the world was reported to engage in insufficient physical activity (PA), which accounted for the death of approximately 3.2 million people every year [17]. With the rapid improvement of technology, people spend less time on household activities and more time watching television, playing video games, or using a mobile phone/computer. Among American children and adults, over half of their waking hours are spent sedentarily [18]. In Europe, people spend about 40% of their leisure time watching TV [18].

There is growing evidence for associations between a sedentary lifestyle and a series of chronic diseases and mortality [18, 19]. Many epidemiological studies have suggested it as an independent risk factor for multiple health outcomes, regardless of the intensity of leisure-time PA levels [18, 20, 21]. Possible associations between a sedentary lifestyle and the development of colorectal, endometrial, ovarian, and prostate cancer have been suggested [19], but few have studies potential associations with lung cancer. Also, previous studies have focused either on occupational sitting [22-24] or leisure-time TV watching [25, 26] in relation to lung cancer risk. Total sitting time daily is suggested to be a better marker, as it reflects sedentary behavior across a range of occupational and leisure-time activities [27]. So far, there have only been two studies on the relationship between total sitting time daily and lung cancer risk [26, 28], and more studies are required.

Even using total sitting time daily as a marker for a sedentary lifestyle, we can only capture the majority of what are considered sedentary behaviors. People who exhibit high levels of prolonged sitting can either have no PA or achieve moderate to high levels of PA during their leisure time. The relationship between PA and lung cancer risk has been studied more extensively. Recent meta-analysis studies concluded that moderate to high levels of PA are associated with a reduced risk of lung cancer in smokers [16, 29, 30]. Like prolonged sitting,

inactivity is also at the lower end of the human movement continuum. Focusing on people who spend a lot of time sitting and are also leisure-time inactive may capture the most sedentary group of people and allow for a better understanding of the possible association between a sedentary lifestyle and lung cancer.

1.4.2 Asthma and asthma symptom control

Asthma is a common chronic lung disease, defined by a history of wheeze, shortness of breath, chest tightness, cough, and sputum production [31]. The most common triggers are tobacco smoking, allergens (e.g., pollens and the house dust mite), exercise, and stress [32]. It affects around 300 million people worldwide [32].

In the clinic, asthma is suspected based on family risk of atopic disease and symptoms. The diagnosis might be confirmed by lung function tests showing variability of bronchial obstruction (bronchodilator responsiveness test) or peak flow diurnal variability, or airway hyperreactivity reported by provocation tests or markers of inflammation (Nitric Oxid) [32]. Another chronic obstructive lung disease, COPD, is a progressive inflammatory disease due to noxious gases or particles, mainly tobacco smoking and is reported as an independent risk factor for lung cancer [33]. The symptoms and obstruction are chronic, but patients may also have exacerbations with increased symptoms. Among people who have long smoking histories or severe asthma, asthma and COPD may co-exist [34]. Before 2000, COPD often was misclassified as asthma in Norway, but stricter diagnostic criteria and different treatment recommendations have improved the validity of the diagnoses [35]. Suspected COPD is confirmed by post-bronchodilator spirometry showing chronic bronchial obstruction. This has been defined as the forced expiratory flow in 1s (FEV1) / forced vital capacity (FVC) < 0.70 [36], but many now recommend a cut-off similar to lower fifth percentile defined by Global lung initiative for Chronic Obstructive Lung Disease (GOLD) 2012 reference values in order to avoid overdiagnosis of COPD among elderly [35].

So far, three meta-analysis studies have investigated the potential association between asthma and lung cancer incidence [37-39]; most included studies were in a case-control study design and suggested a positive association between asthma and lung cancer. Recall bias and reverse causation cannot be excluded in case-control studies. Age and sex-standardized incidence ratio were also used in some studies to present the relative risk for lung cancer in patients with asthma compared with a general population [40-42]. However, there is a lack of adjustment for important confounders in such studies. Prospective cohort studies were limited, and they either focused on lung cancer mortality [43, 44], had a short follow-up duration (<10 years) [45, 46] or had a few lung cancer cases [47, 48]. Smoking is not only the most important risk factor for lung cancer, but also an important trigger for asthma. It can cause more serious asthma symptoms and lead to poorer clinical outcomes [49]. However, the influence of smoking is not well controlled in many studies. One prospective cohort study carried out in 634,039 never smoking women with a follow-up duration of more than 14 years, reported that asthma requiring treatment was associated with an increased incidence of lung cancer [50]; however, asthma was studied among 34 potential risk factors, and a chance finding could not be excluded.

Although asthma is not curable, appropriate treatment can help to control the symptoms, and people with asthma can still have a high quality of life. According to Global Initiative for Asthma (GINA) 2019 guidelines, inhaled corticosteroid (ICS) is the cornerstone in asthma treatment [32]. Through reduced airways inflammation patients experience less symptoms, improved quality of life, reduced exacerbations, and mortality. Previously ICS was only recommended in patients with symptoms or need for reliever therapy more than twice a week [32]. From 2020 GINA has recommended that ICS and bronchodilator should be used for all levels of asthma severity, either regularly or as needed [51]. The adherence to asthma treatment has been rather low, and about half of people with asthma in Europe are reported to

have their asthma either only partly controlled or uncontrolled [52, 53]. Failure to control asthma symptoms may not only increase asthma exacerbations but also lead to functional limitations, higher numbers of comorbidities and incapacity for work [54]. However, little attention has been devoted to the impact of different levels of asthma symptom control on the development of lung cancer.

1.4.3 Body mass index

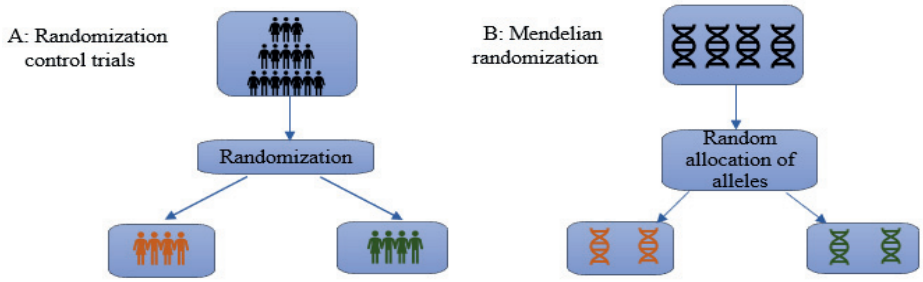
Body mass index (BMI) is calculated as body weight in kilograms per meter squared. Obesity is defined by BMI greater than or equal to 30 kg/m^2, while being overweight is defined by BMI greater than or equal to 25 kg/m^2 and less than 30 kg/m^2 [55]. Since the 1980s, nearly one-third of people worldwide have become overweight or obese [56]. Obesity is becoming a global pandemic in modern society. According to the latest report from the World Health Organization (WHO), in 2016, around 1.9 billion adults aged 18 years and over were overweight, and over 650 million of them were obese [57]. In the US, 68% of adult population are overweight and 39.8 % are obese currently [58].

BMI has been positively associated with several cancers, such as colon, kidney, liver, pancreas, breast (in postmenopausal women), and endometrium cancer [59]. On the contrary, many observational studies have suggested that BMI is inversely associated with the incidence of lung cancer [60-62]. However, the inverse association was mainly found among smokers [63]. The observational evidence has been inconclusive in never smokers [50, 64, 65]. Moreover, the interrelationship between smoking and BMI is complex [66, 67]. Compared with non-smokers, smokers were reported to have a lower BMI [68]. However, it was also reported that heavy smokers tended to have a higher BMI than lighter smokers [68]. Thus, residual confounding by smoking may explain the observed inverse association [50, 65]. In addition, the inverse association may be explained by reverse causation due to weight

loss shortly before the lung cancer diagnosis [69] or competing causes of death due to obesity [70].

Further, BMI varies during the lifespan. The measurements of BMI at one time point in most observational studies may not be sufficient to examine the possible influence of BMI on the development of lung cancer over time [61, 64, 69, 71]. Investigating BMI change between at least two time points would be more informative. Generally, BMI change in relation to the incidence of lung cancer has not been well investigated. One case-control study suggested an inverse dose–response association between BMI gain in adulthood and risk of lung cancer, but more evidence from prospective cohort studies is required [72].

As mentioned above, the methodological pitfalls in traditional observational studies may mask the real relationship between BMI and lung cancer. One approach to solving these problems is to perform Mendelian randomization (MR) analysis. MR analysis is a statistical method that mimics the design of randomized control trials (RCTs) by using genetic variants of the exposure, allowing the inference of causality (Figure 2). Since genetic variants are naturally assigned randomly at conception [73], bias due to reverse causation is avoided, and the influence of residual confounding is less likely. Further, genetically determined BMI reflects BMI across the lifespan and therefore is more accurate than measuring BMI at a single time point [74].



In A, the orange color represents the treated group, while the green color represents the controlled group. In B, the orange color represents effect alleles, while the green color represents control alleles. In both A and B, confounders and measurement errors are equal between groups, and reverse causation is maximally avoided.

Figure 2. Comparison of study design between randomized control trials and Mendelian randomization.

Pleiotropy may exist in conventional univariable MR since some of the genetic variants for BMI may be associated with smoking [75, 76]. Multivariable MR is an extension of univariable MR that can detect the direct causal effects of multiple risk factors simultaneously [77]. Therefore, it can be used to examine the causal effect of BMI on lung cancer incidence, taking the influence of smoking into account [78]. To date, only one such study has been performed, and it suggests an inverse association between BMI and the incidence of lung adenocarcinoma but a positive association between BMI and SCLC [75]. The result on SCLC is consistent with previous univariable MR studies [74, 79]. However, the inverse association with adenocarcinoma differs from previous univariable MR studies, which suggested no association. Thus, more multivariable MR study is needed to clarify the causal association between BMI and the incidence of lung cancer, in particular with the lung adenocarcinoma.

1.5 Possible mechanisms linking risk factors other than smoking and lung cancer

There are several plausible pathways that may explain the possible association of a sedentary lifestyle or asthma with lung cancer. Some animal studies have shown that a lack of activity

might suppress lipoprotein lipase activity in skeletal muscles and reduce glucose uptake [80, 81]. Both are related to metabolic disorders, which are a known risk factor for several malignancies [81, 82]. In addition, some pre-clinical studies suggest that weight-bearing skeletal muscles are not highly engaged during inactivity [83-85]. This may alter the anti-cancer responses of myokines in skeletal muscles and activate inflammatory pathways that are important for cancer development [83-85].

Asthma is characterized by chronic airway inflammation [38, 86]. The inflammatory response in the lung may be associated with the lung carcinogenic process through elevated levels of free radicals and reduced levels of antioxidants [87]. This may lead to increased DNA damage and mutations [38], and permanent abnormality of the airways [88].

With respect to the possible mechanisms behind the influence of BMI, a paradox may exist: some adipokines from adipose tissue may produce tumor-suppressing effects [89], while others affect the progress of carcinogenesis through irregular immunomodulation or chronic inflammation [90, 91]. It is also reported that high insulin resistance related to obesity may contribute to the lung carcinogenesis [90, 92]. Thus, more studies are required to clarify the association between BMI and lung cancer.

2 STUDY AIMS

2.1 Overall aim

The overall aim of this thesis was to examine the potential associations of some other risk factors than smoking (i.e., a sedentary lifestyle, asthma, and BMI) with the incidence of lung cancer and its histologic types in a large homogeneous population of Norwegian adults with a long follow-up duration.

2.2 Specific aims of Studies 1–3

Study 1: The aims of this study were to prospectively explore 1) the relationship between total sitting time daily and the incidence of lung cancer overall and its histologic types and 2) to investigate how the different combinations of total sitting time daily and leisure-time PA were associated with the incidence of lung cancer overall and its subtypes.

Study 2: The aim of this study was to prospectively explore the potential associations of asthma overall as well as asthma status and symptom control with the incidence of lung cancer overall and its subtypes.

Study 3: The aim of this study was to first explore the potential influences of BMI and BMI change on the incidence of lung cancer overall and its subtypes using an observational approach. The possible causal associations were further examined using a one-sample multivariable MR approach after genetically controlling for smoking.

3 MATERIALS AND METHODS

3.1 The HUNT Study

The Trøndelag Health Study (HUNT), formerly known as the Nord-Trøndelag Health Study, is a large population-based health study conducted mainly in the north area of Trøndelag, Norway (Figure 3). The north area of Trøndelag was one of 19 Norwegian counties until 2018, when it was merged with the county of Sør-Trøndelag to form the Trøndelag county [93]. It includes 24 administrative municipalities and has two local hospitals, one in Levanger and the other in Namsos. More than 97% of Trøndelag's inhabitants are Caucasian, which makes the HUNT population largely homogenous. The adult population had also been relatively stable between the first (125,835 in 1984) and the third HUNT survey (128,694 in 2006). There is a low net migration out of the north area of Trøndelag (0.3% per year on average from HUNT1 to HUNT3) [93]. The region's socio-demographic characteristics as well as its mortality and morbidity data also mirror those of Norway. Thus, the HUNT Study is a suitable source of data for performing prospective cohort studies.

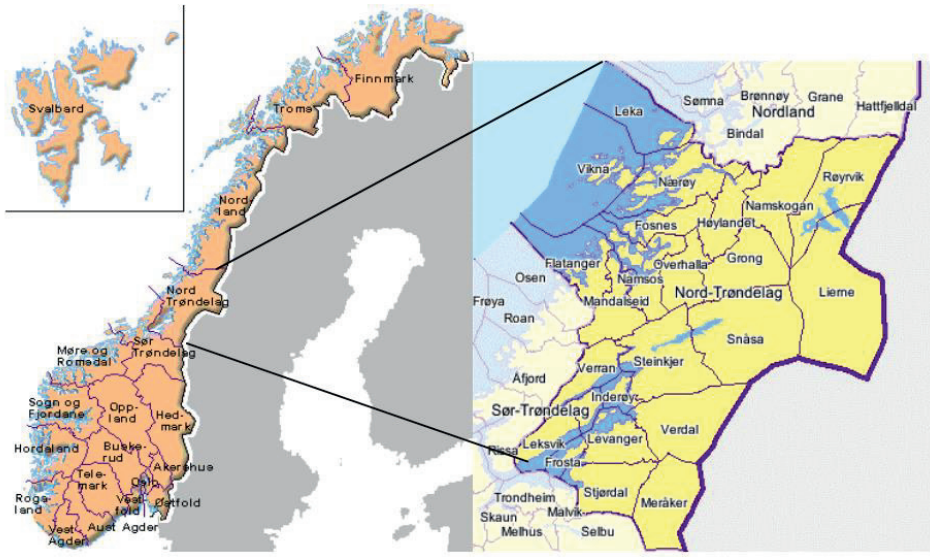


Figure 3. Norway and the north area of Trøndelag. Adapted from Holmen et al. (2003) Norwegian Journal of Epidemiology [93].

The HUNT Study currently consists of four consecutive surveys: HUNT1 (1984–1986), HUNT2 (1995–1997), HUNT3 (2006–2008), and HUNT4 (2017–2019) [93, 94]. It is a collaborative project of the HUNT Research Centre of Faculty of Medicine and Health Science, Norwegian University of Science and Technology, the Norwegian Institute of Public Health, and Trøndelag County Council. For the adult part of each survey, all adults aged 20 years or older living in the county were invited to complete extensive health and lifestyle questionnaires, and to undergo a clinical examination [94]. In total, the HUNT Research Centre has so far created a database for nearly 145,000 people. Most of the participants took part in more than one survey. In HUNT2-4, DNA was extracted from blood samples, and genome-wide association study analyses were performed. In addition, the HUNT Study includes some sub-studies, for example the Lung Study, which have selected participants to undergo further examinations and interviews and answer specific questionnaires [94]. More information on HUNT questionnaires and selection procedures can be found at http://www.ntnu.edu/hunt.

3.2 Study populations

3.2.1 HUNT2 and HUNT2 Lung Study

HUNT2 is the baseline for all three studies included in this thesis. HUNT2 took place between 1995–1997. A total of 94,194 adults aged 20 years or older were invited to participate. Among them, 65,229 (70%) participants completed self-administered questionnaires, and were invited to further questionnaires, clinical measurements, and blood samples. All participants were asked to fill in a general questionnaire at home (step one) and then handed it to a health professional when they took the clinical examination. The participants received a second questionnaire (step two) with more detailed questions about their health and lifestyle during the clinical examination with a prepaid envelope. The questionnaire was completely at home and returned by mail.

Further, the HUNT2 Lung Study selected a 5% random sample of the total HUNT2 participants as well as an asthma symptom group and collected more detailed information on history of asthma diagnosis, attacks of wheezing or breathlessness during the last 12 months and use of asthma medication. Lung function was measured by a spirometry test [94]. All participants in the Lung Study performed pre-bronchodilator spirometry tests. Subsequently, those participants from the five urban municipalities who had an airflow limitation defined by a pre-bronchodilator spirometry result of $FEV_1/FVC < 0.75$ or $FEV_1 < 80\%$ and all the participants from the 19 rural municipalities were further invited to perform pre-bronchodilator and post- bronchodilator spirometry tests.

3.2.2 Study design

The first two studies are prospective cohort studies with two measurement points (baseline and follow-up). Information on exposures (i.e., total sitting time daily, its combination with leisure-time PA, asthma, asthma symptom control, and important covariates) was collected from people free from lung cancer in HUNT2 and they were followed up over time to observe the incidence of lung cancer. Using unique 11-digit personal identification numbers, participants' information from HUNT2 was linked to the Cancer Registry of Norway [9]. The follow-up period was 18.3 years for the first study and was 21.1 years for the second and third studies with updated information on lung cancer cases from the Cancer Registry of Norway.

In the third study, both BMI in HUNT2 and BMI change from HUNT1 to HUNT2 were used as exposure measures. The multivariable MR method was used to study the possible causal association between BMI and lung cancer incidence, genetically controlling for smoking.

Study 1: In this study, every participant was followed up from the date of participation in HUNT2 until the date of first diagnosis of lung cancer, the date of death, the date of emigration from Norway, or the end of follow-up on December 31, 2014, whichever came

first. We first excluded 6159 participants with previous cancer. This left 59,070 participants who self-reported no cancer. We then further excluded 13,260 participants who did not have information on total sitting time daily, leaving 45,810 cancer-free participants with complete information on total sitting time daily in the main cohort. We also investigated the combination of total sitting time daily and leisure-time PA in relation to lung cancer incidence in a sub-cohort of 33,793 participants who provided complete information on leisure-time PA. A flow chart of the study population for Study 1 (Paper I) is given in Figure 4.

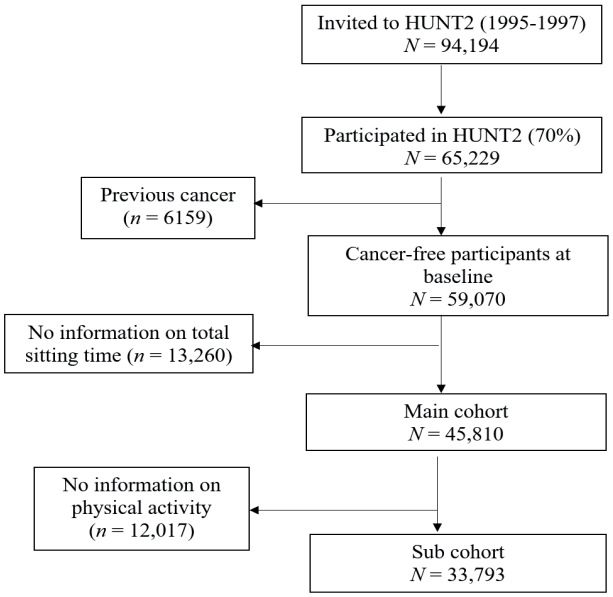


Figure 4. Flow chart of the study population in Study 1 (Paper I).

Study 2: In this study, every participant was followed up from the date of participation in HUNT2 until the date of first diagnosis of lung cancer, the date of death, the date of emigration from Norway, or the end of follow-up on December 31, 2017, whichever came first. Information on asthma symptoms and lung function was collected from the HUNT2 Lung Study.

We first excluded 2053 participants with cancer at or before the baseline based on information from the Cancer Registry of Norway. Additionally, we excluded 71 participants without information on asthma. To minimize the influence of COPD, we further excluded 312 adults who had possible COPD based on the GOLD definition [95]: 1) reported doctor-diagnosed COPD, 2) a post-bronchodilator of FEV1/FVC <0.7 and 3) ever smokers at baseline. Reported doctor-diagnosed COPD was defined based on the question "Have you been diagnosed as having chronic bronchitis or emphysema by a doctor?". Lung function was measured by spirometry in the HUNT2 Lung Study. The post- bronchodilator fixed ratio (FEV1/FVC <0.7) is the most used spirometry criterion for diagnosing COPD in the literature [95]. Finally, 62,791 participants were left for the analyses. A flow chart of the study population in Study 2 (Paper II) is given in Figure 5.

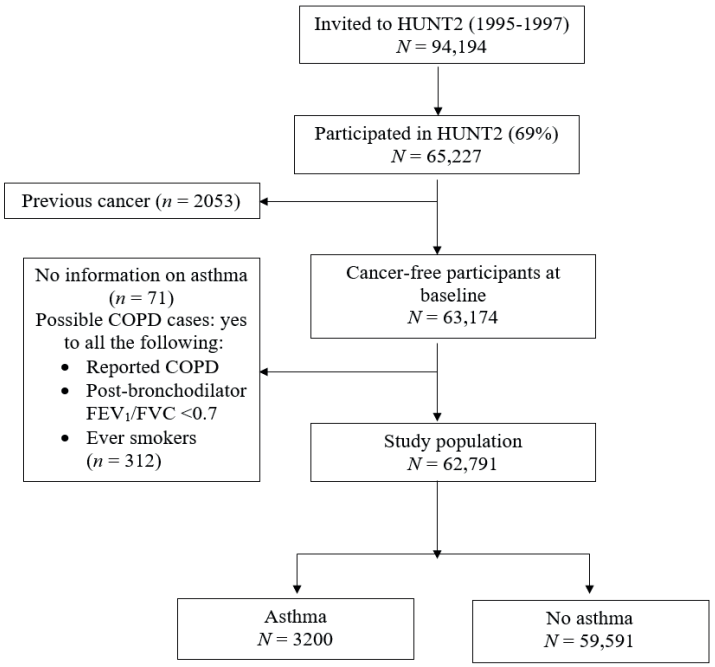


Figure 5. Flow chart of study population in Study 2 (Paper II).

Study 3: In this study, we first excluded 2053 participants with cancer at or before the baseline based on information from the Cancer Registry of Norway. Additional 721 participants with missing information on BMI in HUNT2 were also excluded, leaving 62,453 adults for the analysis of the relationship between BMI in HUNT2 and lung cancer incidence. Next, we excluded 18,060 participants with missing information on BMI in HUNT1, leaving 44,393 participants for the analysis of BMI change from HUNT1 to HUNT2 and lung cancer incidence. Among the 18,060 participants who had missing information on BMI in HUNT1, about 50% of them were aged less than 20 years and were not eligible for HUNT1. For the multivariable MR analysis, 7942 participants without information on genetic variants for BMI in HUNT2 were excluded, leaving a total of 54,511 participants for the analysis. A flow chart of the study population in Study 3 (Paper III) is given in Figure 6.

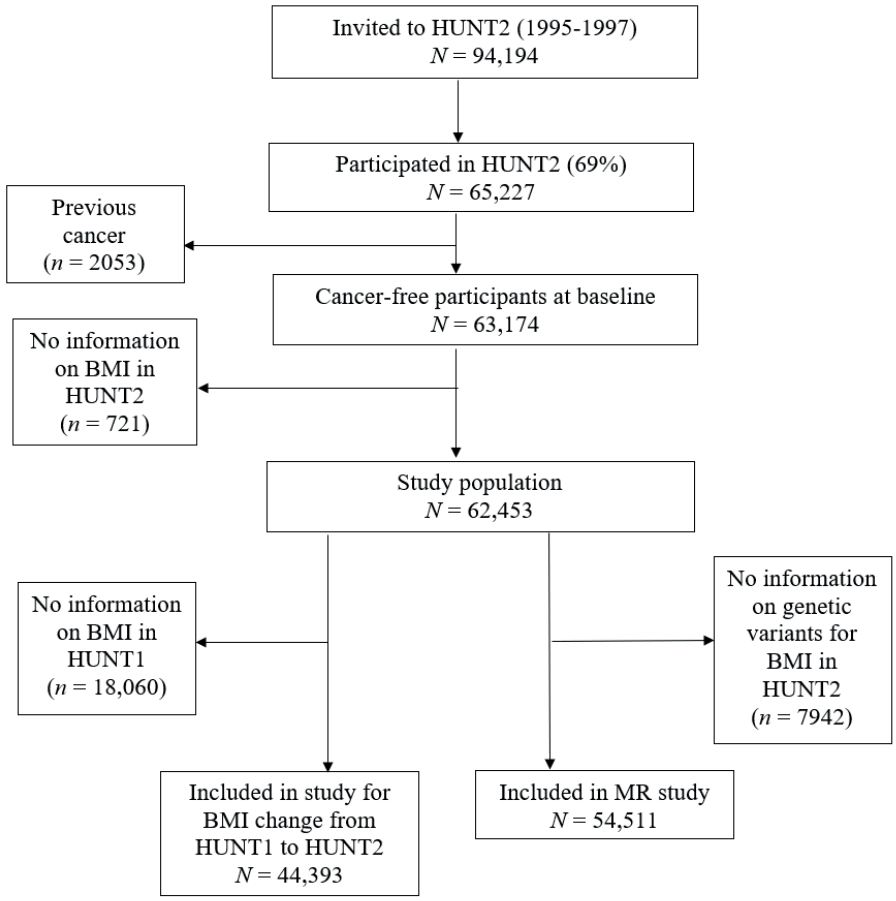


Figure 6. Flow chart of study population in Study 3 (Paper III).

3.3 Study variables

3.3.1 Measurements of exposures

Total sitting time daily: Total sitting time daily was assessed by the following question: "How many hours do you usually spend in the sitting position during a 24-hour period? (Work, meals, television and car etc.)". The total number of hours reported by participants (as a positive integer) were grouped total into three categories (0–4 hours, 5–7 hours and ≥8 hours) based on similar cutoff criteria in earlier HUNT studies [96, 97] and two meta-analysis studies [98, 99].

Occupational activity: Occupational activity was used as an alternative marker for a sedentary lifestyle in Study 1 and regarded as a potential confounder in Study 3. It was measured by the question: "How would you describe your work?", with participants able to choose from four response options. Responses were categorized as mostly sedentary work, much walking or lifting (two response options collapsed into one category), heavy physical work, or "unknown" (due to missing information).

Leisure-time PA: Leisure-time PA was measured by the question "How much of your leisure time have you been physically active per week during the last year?". Participants were asked to specify average hours of light (no sweating or not being out of breath) and hard PA (sweating or out of breath) per week, with the following response options for each intensity: none, <1 hour, 1–2 hours, and ≥3 hours (reported as a positive integer). We classified the participants' PA levels as inactive (no activity or ≤2 hours of light activity), low (≥3 hours of light activity only or ≤2 hours of light activity and <1 hour of hard activity), moderate (≥3 hours of light activity and <1 hour of hard activity or 1–2 hours of hard activity regardless of light activity) and high (≥3 hours of hard activity regardless of light activity). This classification demonstrated a dose–response relationship with mortality [100]. Based on the new variable, people's PA levels from low to high were collapsed into one category and defined as physically active. People who were physically inactive reported no activity or ≤2 hours of light activity.

Combined categories of total sitting time daily and leisure-time PA: Based on information about total sitting time daily and leisure-time PA, we defined four combined categories: 1) total sitting time of <8 hours daily and physically active, 2) total sitting time of <8 hours daily and physically inactive, 3) total sitting time of ≥8 hours daily and physically active and 4) total sitting time of ≥8 hours daily and physically inactive.

Asthma: Asthma was defined by affirmative answers to the following two questions: "Do you have, or have you had asthma?" and "Have you been diagnosed as having asthma by a doctor?". Asthma status was further categorized into active and non-active asthma. Participants were considered to have active asthma if they confirmed symptoms of wheezing or reported using asthma medication in the last 12 months.

Asthma symptom control: The following four items were used to describe the level of asthma symptom control: 1) daytime symptoms more than twice weekly, 2) any night awakening, 3) need for reliever medications more than twice weekly or 4) any activity limitation based on GINA Global Strategy for Asthma Management and Prevention guidelines [32]. The questions in the HUNT2 Lung Study matched the GINA guidelines (Table 1). According to the GINA guidelines, controlled asthma refers to asthma without any of the above four indicators, partly controlled asthma refers to asthma with 1–2 of the indicators and uncontrolled asthma is asthma with 3–4 of the indicators. We initially classified levels of asthma symptom control as controlled, partly controlled, uncontrolled and unknown. Partly controlled and uncontrolled were then collapsed into one "partially controlled" category due to the uncontrolled category having only a few lung cancer cases.

Table 1. Comparison of assessment of asthma symptom control between the GINA assessment and the questions in HUNT2

	GINA assessment	**HUNT2**
1	**Daytime symptoms**	
	In the past 4 weeks, has the patient had daytime asthma symptoms more than twice/week?	Have you at any time in the last 12 months been short of breath when resting during the day? (Yes/No)
	Controlled: ≤2 times/week	Controlled: No
2	**Night awakening**	
	In the past 4 weeks, has the patient had: Any night waking due to asthma?	Have you woken up with a feeling of tightness in your chest at any time in the last 12 months? (Yes/No)
	Controlled: 0 times/week	Controlled: No
3	**Reliever medication use**	
	In the past 4 weeks, has the patient had: Reliever for asthma needed more than twice/week?	If you currently use asthma medicines, have you used short-acting beta2 agonists in the last month? (Never, once or less often a week, many times a week, daily)
	Controlled: ≤2 times/week	Controlled: ≤ once or less often a week
4	**Activity limitation**	
	In the past 4 weeks, has the patient had: Any activity limitation due to asthma?	How much do your respiratory problems (dyspnea) affect your daily activities? (Not at all, a little, much, very much)
	Controlled: 0 times/week	Controlled: ≤ a little

Abbreviations: GINA: Global Initiative for Asthma; HUNT: The Trøndelag Health Study

BMI and BMI change: Weight and height in HUNT1 and HUNT2 were measured by health professionals at clinical examinations. Height was measured to the nearest centimeter and weight to the nearest 0.5 kg. BMI was calculated as weight in kilograms divided by height squared in meters (kg/m^2) and was initially grouped into: <18.5, 18.5–24.9, 25.0–29.9, 30.0–34.9 and ≥35.0 kg/m^2 according to the recommendations of WHO [101]. Due to limited lung cancer cases in the BMI categories <18.5 and ≥35.0 kg/m^2, the BMI categories were collapsed into three groups such as: <25.0, 25.0–29.9 and ≥30.0 kg/m^2. BMI changes from

HUNT1 to HUNT2 were categorized into quartiles: 1st (-21.3–0.5), 2nd (0.6–1.7), 3rd (1.8–3.1) and 4th (3.2–18.6) in kg/m^2.

Genetically determined BMI: DNA samples were extracted from blood samples collected during HUNT2 and stored in the HUNT Biobank. Genome-wide genotyping and imputation were carried out for all participants in HUNT2 with sample and variant quality control by using Illumina Humina HumanCoreExome arrays [102]. Seventy-seven single nucleotide polymorphisms (SNPs) were suggested as candidate instrumental variables for BMI with a P value of $<5 \times 10^{-8}$ based on European sex-combined analyses in a genome-wide association study by the Genetic Investigation of Anthropometric Traits (GIANT) consortium [103]. Information on 2 SNPs (rs12016871 and rs2033732) was missing in the HUNT data since they did not pass imputation quality control, leaving 75 BMI-associated SNPs for our analysis. We then split the remaining 75 BMI-associated SNPs into two groups [75, 78] : 1) 61 SNPs that only affected BMI (BMI-only SNPs) and 2) 14 SNPs that affected both BMI and smoking (BMI & smoking SNPs). The 14 BMI & smoking SNPs were identified based on a P value <0.05 for the association between each of the 75 BMI-associated SNPs and smoking in our study population.

3.3.2 Ascertaining lung cancer

The International Classification of Diseases version 10 (ICD-10) codes used for registration of lung cancer are C33-C34 [104]. Histologic types were classified according to the International Classification of Diseases of Oncology (ICD-O) [105]. The Cancer Registry of Norway, founded in 1951, includes records of all cancer cases nationally from January 1952. In addition, the Cancer Registry of Norway matches all information on new cancer cases with data from death certificates and the Norwegian Patient Registry. Data from the Cancer Registry of Norway are reasonably accurate and complete [106].

3.3.3 Other important variables

Age, sex, smoking and passive smoking, alcohol consumption, education, economic difficulties, and family history of cancer were included *a priori* as potential confounders in all the studies. Smoking was classified into seven categories based on detailed information of smoking status and pack-year: as never, former (≤10.0, 10.1–20.0, and >20.1 pack-years), and current smokers (≤10.0, 10.1–20.0, and >20.1 pack-years). Other variables were categorized as follows: passive smoking (never, only in childhood, only in adulthood, and both), alcohol consumption (never, 1–4 times/month, and ≥5 times/month), education (reported as a positive integer, categorized into three categories in study 1 and 2: <10, 10–12, and ≥13 years and simplified into two categories in study 3: <10 and ≥10 years), economic difficulty ("During the last year, has it at any time been difficult to meet the cost of food, transportation, housing and such?" [Yes or no]) and family history of cancer ("Is there any family member such as father, mother, siblings who reported cancer?" [Yes or no]).

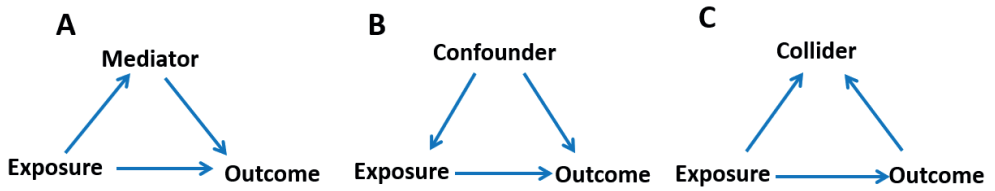
Chronic bronchitis ("Have you had a cough with phlegm for periods of at least three months during each of the last two years?" [Yes or no]) was included in Study 1 as a proxy for COPD. Participants were considered to have allergic rhinitis if they reported having allergic rhinitis in combination with allergy medication use or allergic symptoms around pollen or pets. This variable was included in Study 2. Use of ICS was included in Study 2's sensitivity analysis to evaluate the potential influence of ICS on the risk of lung cancer. People with regular ICS use were those who answered "yes" to the question "Have you ever regularly used medicines like Becotide, Flutide, Pulmicort or Viarox" in the HUNT2 Lung Study. In Study 3, self-reported COPD was determined by a positive response to the question "Have you been diagnosed as having chronic bronchitis or emphysema by a doctor?". An "unknown" category was created for each of above variables with missing information.

3.4 Statistical analysis methodology

Directed acyclic graphs (DAGs) were used to visualize hypothetical relations among the variables of interest in the present thesis. Survival analysis and MR analysis were the main statistical approaches used for the thesis studies to determine the potential associations between risk factors other than smoking and the incidence of lung cancer and its subtypes. Sensitivity analyses were conducted to ensure the robustness of main findings. All statistical analyses were performed with STATA/SE 14.2, 15.1, or 16.1 (College Station, TX, USA). We also used the package "MendelianRandomization" from R to perform the multivariable MR in Study 3. These statistical concepts and analysis methods are described in detail in the following subsections.

3.4.1 Directed acyclic graphs

DAGs were drawn based on previous knowledge of the possible relations between the exposures of interest, covariates, and each outcome in the three studies of the thesis [107-109]. In a DAG, a path refers to the sequence of arrows connecting two variables, irrespective of the direction of the arrows [110]. An exposure and the outcome can be connected either directly or indirectly. There are three possible ways they can be connected indirectly (Figure 7). If all the arrows in the path are in the same direction, the covariate is a mediator on the causal pathway from exposure to outcome. If the arrows point from the covariate toward both the exposure and outcome, the covariate is a common cause of both the exposure and outcome—it is a potential confounder. When the arrows point from the exposure and outcome toward the covariate, the covariate is a common effect of both the exposure and outcome, and it is regarded as a collider.



A: Covariate is a mediator; B: Covariate is a confounder; C: Covariate is a collider.

Figure 7. Directed acyclic graphs show the possible underlying relationships between covariates and their associated exposures and outcome.

Adjusting for a mediator or collider creates bias and should be avoided [110]. However, it is important to adjust for a confounder to obtain a more valid estimate for the effect of an exposure on the outcome. DAGs are usually used as an established framework to properly identify all necessary confounders for the analysis of causal inference in epidemiology [111].

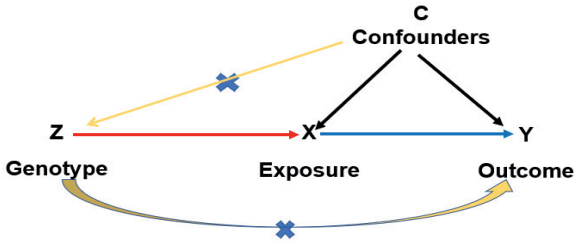
3.4.2 Survival analysis

We used survival analysis to examine the prospective associations between exposures and the incidence of lung cancer and its histologic types. Cox proportional hazard models were used to present crude and adjusted hazard ratios (HR) with 95% confidence intervals (CI). A HR, which should be constant over time, represents the relative risk of an event in one category of exposure compared with the risk in the reference category after adjusting for covariates [112]. A HR equal to 1 means there is no difference between the compared categories of exposure. A HR over 1 suggests that the hazard associated with the event of interest has increased with the exposure of interest, while an HR between 0 and 1 suggests that the event is associated with a reduced hazard. It is necessary to assess the proportional hazard assumption using both statistical tests and graphs based on Schoenfeld residuals. In this thesis, age was used as the time scale, with entry and exit time defined as the subject's age at recruitment and age at lung cancer diagnosis or censoring, respectively. Variables that were suggested against the

proportional hazard assumption were treated as covariates with time-varying effect, and the *tvc* option of the *stcox* command in Stata was used to model the non-proportional hazards.

3.4.3 Mendelian randomization

MR is an important complement to traditional observational methods, as it uses genetic variants as instruments for the exposure of interest to examine causal associations. The advantage of MR is that genetic variants are naturally assigned randomly before conception [73]. Bias due to reverse causation is thus avoided, and the influence of residual confounding is less likely. Conventionally, in univariate MR, genetic variants are used as instruments for only one risk factor (Figure 8). There are three key assumptions for a valid univariable MR analysis: 1) the "relevance" assumption, which suggests that the genetic variants are associated with the exposure; 2) the "independence" assumption, which suggests that there are no associations between the genetic variants and the potential confounders; and 3) the "exclusion" assumption, which suggests there is an association between the genetic variants and the outcome only through the exposure. If any of the three assumptions are violated, for example, by pleiotropy, the causal estimates may be biased, and erroneous conclusions may be drawn [113].



Valid instrumental variables (Z) are defined by three assumptions in univariable MR: 1) relevance (Z is associated with exposure X); 2) independence (no common cause C is between Z and the outcome Y); and 3) exclusion (Z is only associated with the outcome Y through exposure X).

Figure 8. The three assumptions in univariable MR.

Multivariable MR is an extension of univariable MR that can better account for pleiotropy. In multivariate MR, it is possible to detect the direct causal effects of multiple risk factors simultaneously since the genetic variants are used to explain multiple exposures. As such, the "independence" and "exclusion" assumptions are extended accordingly [77].

3.4.4 Main analyses

Study 1: Based on results from the likelihood ratio test, non-linearity was suggested for total sitting time daily and the incidence of lung cancer overall (P = 0.02). Thus, the three categories (0–4 hours, 5–7 hours, and ≥8 hours) of total sitting time were applied to represent total sitting time daily in the main analysis. Total sitting time was also categorized into tertiles to test the robustness of the results from the main analysis. The potential influence of total sitting time (sitting for ≥8 hours) on the incidence of lung cancer was further studied by combining it with leisure-time PA (active or inactive). We categorized lung cancer into two major subtypes, SCLC and NSCLC, as the statistical power was reduced (with wider CIs) for further categorization into adenocarcinoma and squamous cell lung cancer in our sub-cohort when we studied total sitting time daily combined with leisure-time PA. The potential confounders in the adjusted models were sex, BMI, smoking, passive smoking, alcohol consumption, education, economic difficulties, family history of cancer, and chronic bronchitis (as a proxy for COPD).

In Paper I, we tested the proportional hazard assumption for exposures, and the results suggested that the hazard functions of both total sitting time daily and its combination with leisure-time PA were proportional. We deemed that the violation of the proportional hazard assumption for some covariates would not have major influence on our results, so all covariates were treated as covariates without time-varying effect in Paper I. We further tested the proportional hazards assumptions for all important covariates when preparing the thesis. We found that the proportional hazard assumption was not met for smoking and economic

difficulties in relation to lung cancer overall. No covariates showed evidence against the proportional hazard assumption for either SCLC or NSCLC. Treating smoking and economic difficulties as covariates with time-varying effect showed similar results that were presented in Paper I.

Study 2: In Study 2, sex, BMI, smoking, passive smoking, alcohol consumption, leisure-time PA, total sitting time daily, education, economic difficulties, family history of cancer, and allergic rhinitis were included as confounders in the adjusted models for the associations of asthma overall, asthma status, asthma symptom control with lung cancer incidence. We performed analyses for both lung cancer overall and its subtypes as SCLC and NSCLC separately. We only presented the results based on lung cancer overall in Paper II. The analyses for SCLC and NSCLC are added as additional results in this thesis. Testing the proportional hazard assumption revealed that the covariates of sex, smoking and economic difficulties did not meet the proportional hazard assumption, and were treated as covariates with time-varying effect in the adjusted models for lung cancer overall. Allergic rhinitis was included in the adjusted models as a potential confounder because it closely associated with asthma and is inversely associated with lung cancer risk [114].

Study 3: In the observational part of Study 3, BMI was regarded as both a continuous and categorized variable. We merged the BMI categories into <25.0, 25.0–29.9 and ≥30.0 kg/m^2 in the main analysis since there were limited lung cancer cases for the BMI categories of <18.5 (n = 11) and ≥35 (n = 30) kg/m^2. As a continuous variable, the likelihood ratio test showed that BMI did not violate the assumption of linearity (P = 0.57). Sex, smoking, passive smoking, leisure-time PA, total sitting time daily, education, economic difficulties, family history of cancer, and self-reported COPD were adjusted in the main models. Asthma, alcohol consumption, and occupational activity were adjusted in the additionally adjusted models. For the associations between BMI and the incidence of lung cancer and its subtypes,

the covariates with non-proportional hazards were sex, smoking, and economic difficulties for lung cancer overall; smoking, family history of cancer, economic difficulties, and leisure-time PA for SCLC; sex, smoking, economic difficulties, and leisure-time PA for adenocarcinoma; and education for squamous cell lung cancer. For the associations between BMI change and the incidence of lung cancer overall and for its subtypes, the covariates with non-proportional hazards were sex, smoking, and economic difficulties for lung cancer overall; smoking for SCLC; sex, smoking, and economic difficulties for adenocarcinoma; and education for squamous cell lung cancer.

In the MR part of Study 3, we performed multivariable MR analysis to assess the potential causal association between genetically predicted BMI per 1 kg/m^2 increase and the incidence of lung cancer overall and of its subtypes, accounting for the influence of smoking genetically. All 75 BMI-associated SNPs (61 BMI-only SNPs and 14 BMI & smoking SNPs) were used as instrument variables, with the r^2 measure of linkage disequilibrium among the instruments <0.01 at a 10-MB window [103]. BMI and smoking in HUNT2 were both regarded as exposures in the multivariable MR. We obtained estimates for the associations between SNPs and BMI, between SNPs and smoking (regarded as an ordinal variable) and between SNPs and lung cancer from the same individuals and applied two-sample MR methods, such as the inverse variance weighted (IVW) and MR-Egger methods. Both methods can be used in a one-sample setting [115, 116]. We calculated the regression coefficients and their standard errors for the SNPs–BMI associations, with adjustment for sex, age and age-squared [103]. No adjustments were made for the SNPs–smoking and SNPs–lung cancer associations since no associations other than smoking were identified between the 75 SNPs and the other confounders. The Sanderson-Windmeijer conditional F-statistic was used to estimate the strength of the instruments for BMI conditional on smoking [117]. In addition, Cochran's Q test was used in both the IVW and MR-Egger to detect the

heterogeneity of the ratio estimates. The intercept test of the MR-Egger was also used to assess the possibility of horizontal pleiotropy [113], as indicated by a non-zero intercept and a P value of <0.05. The outlier SNPs in the multivariable IVW and MR-Egger regression methods were identified using the MR-Pleiotropy Residual Sum and Outlier (MR-PRESSO) [118] method.

3.4.5 Sensitivity analyses

Observational parts

Misclassification of exposures: In Study 1, we first conducted a sensitivity analysis to test the robustness of our results on the association between total sitting time daily and lung cancer incidence by using occupational inactivity (mostly sedentary work) as exposure. We then combined the groups of low physical activity and inactivity (no activity, any light activity only, or light activity ≤2 hours and hard activity <1 hour weekly) and repeated the analysis. This allowed us to test whether the original category of a combined sitting time of ≥8 hours and physical inactivity (the original definition of inactivity referred to no activity or ≤2 hours of light activity) captured the most sedentary individuals.

In Study 2, we used two means to reduce possible misclassification of COPD as asthma: we excluded participants with asthma who had smoked 10 pack-years or more and were older than 40 years at the time of asthma diagnosis and were with a) post-bronchodilator FEV_1/FVC <0.7 (fixed ratio criterion) (n = 104) or b) post- bronchodilator FEV_1/FVC z score <-1.64 (lower limit of normal criterion) (n = 75). The fixed ratio criterion is the most often used approach to define airflow limitation, whereas the lower limit of normal criterion overcomes the overestimation of COPD among the elderly [35, 95].

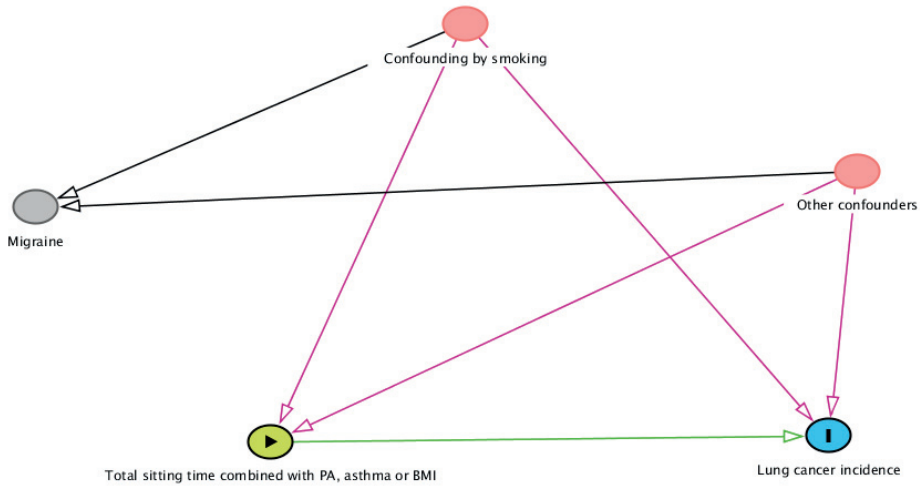
Reverse causality: To address reverse causation, exclusion of the first five years of follow-up has been performed in many studies [119, 120]. However, sometimes, exclusion of the first five years may lead to the loss of too many cases of the outcome and weaken the study's

power. Thus, some studies have preferred exclusion of the first four years [50] or three years [121, 122]. In Studies 1 and 2, when excluding the first five years of follow-up (n = 32,884 in Study 1, n = 59,944 in Study 2), the statistical power was notably reduced, and no clear conclusion could be drawn. Therefore, we instead excluded only the first three years to a better precision of the results. In Study 3, we were able to exclude the first five years of follow-up (n = 59,711) without substantially affecting the study's statistical power.

Residual confounding by smoking: In Study 1, we conducted a complete case analysis for individuals with complete information on smoking (n = 31,907) to minimize any residual confounding from smoking in a sub-cohort and smoking that might affect the association estimates between combined total sitting time daily and leisure-time PA and lung cancer incidence. The results were presented in Paper I. Additionally, in the thesis, we conducted multivariable chained imputation with fully conditional specification (m = 10 imputed datasets) for all variables including smoking, based on the assumption of missing at random [123]. In Studies 2 and 3, multivariable chained imputation was performed, and the results were presented in Papers II and III.

Further, we performed analysis with a negative control exposure study, using migraine as an alternative exposure to further address residual confounding by smoking in all three studies (negative control exposure analysis was performed as sensitivity analysis in Papers II and III, and as an additional analysis in Study 1 at the time of writing the thesis) [124]. Previous studies have suggested that migraine is associated with smoking [125-127] but not with lung cancer. We expected to observe a null association between migraine and lung cancer after adjustment for smoking, which would indicate that the observed associations were less likely to be biased by residual confounding from smoking (Figure 9). Participants with migraine were those who answered yes to the question "Have you suffered from headaches during the last 12 months?" and specified the type of headache as "migraine." In this analysis, we first

adjusted for smoking only and then adjusted for the same confounders as in the primary studies.



This DAG was created using DAGitty V 3.0 [128] (http://www.dagitty.net/dags.html).

Figure 9. Directed acyclic graph for exposures (total sitting time combined with PA, asthma, or BMI), negative control exposure (migraine), and outcome (lung cancer incidence).

Competing risk of death: Death is a significant and related competing risk when using survival analysis to study cancer incidence [129]. To deal with possible competing risk due to death, a competing risk analysis based on the Fine-Gray model was used in this thesis. Competing risk analysis was performed as sensitivity analysis in Papers II and III and as an additional analysis in the thesis for Study 1 [130].

MR parts

Univariable MR: In Study 3, we applied univariable MR methods based on the 61 BMI-only SNPs as sensitivity analyses to test the robustness of the results from the multivariable MR analyses. We used a two-stage method based on a weighted BMI genetic risk score (GRS) to reduce weak instrument bias [131]. GRS based on the 61 BMI-only SNPs was found to explain 2% of the variance in BMI in HUNT2, corresponding to an F-statistic of 905 [132]. In addition, analyses were performed using the IVW and MR-Egger methods, based on summarized data of the 61 individual BMI-only SNPs. The outlier SNPs in the IVW and MR-Egger regression methods were identified using the MR-PRESSO method [118].

3.5 Ethics

The project was approved by the Regional Committee for Medical and Health Research Ethics of South-East Norway (2015/78/REC South-East). All participants provided informed written consent for their participation in HUNT.

4 RESULTS

4.1 Study 1

Among the 45,810 participants in the main cohort, 549 participants developed lung cancer during a median follow-up time of 18.3 years. Compared with those sitting for 5–7 hours daily, the participants sitting for ≥8 hours were more likely to be male, alcohol drinkers, and less physically active, and to have a higher education level and sedentary work. Participants who sat for ≥8 hours daily and were physically inactive during their leisure time were more frequently smokers and non-drinkers, more often had sedentary work, had a higher BMI and lower education compared with participants who sat for <8 hours daily and were physically active. Compared with the original cancer-free population, the participants in the main cohort had a higher percentage of family history of cancer, and the participants in the sub-cohort were relatively younger.

Unlike in the published article (Paper I), we here treated smoking and economic difficulties as covariates with time-varying effects in the Cox models (Table 2 and Figure 10), and the results were similar to those obtained in Paper I. Total sitting time daily (classified by category) was not shown to be associated with incidence of lung cancer overall, SCLC, or NSCLC in the main cohort after adjusting for smoking (based on detailed information on smoking status and pack-years), leisure-time PA, and other important confounders (Table 2). Similar results were found for total sitting time classified into tertiles (data not shown). Due to limited lung cancer cases among never smokers (n = 26), we were not able to perform analysis with this group and only performed analysis with the ever smoker group. The results were similar to our main results.

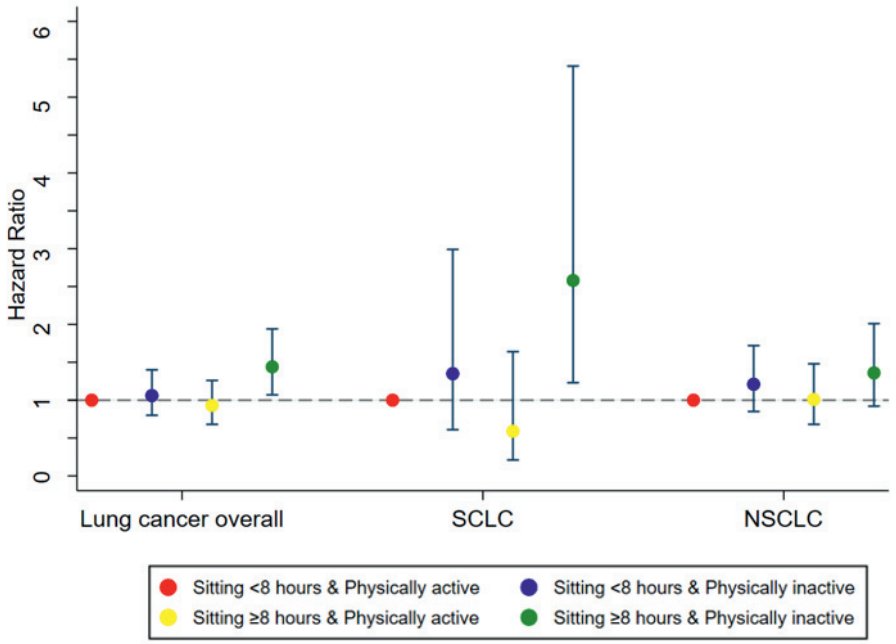
Table 2. The association of total sitting time daily with the incidence of lung cancer overall and its major histologic types. Data from the HUNT Study, 1995–97 to 2014 (N = 45,810)

		Cases	Adjusted[1] HR	95% CI
LC overall	Sitting 0–4 hours	185	1.00	Reference
	Sitting 5–7 hours	165	0.85	0.69–1.05
	Sitting ≥8 hours	199	1.12	0.91–1.38
SCLC	Sitting 0–4 hours	25	1.00	Reference
	Sitting 5–7 hours	20	0.73	0.40–1.31
	Sitting ≥8 hours	31	1.18	0.69–2.03
NSCLC	Sitting 0–4 hours	117	1.00	Reference
	Sitting 5–7 hours	97	0.76	0.58–1.00
	Sitting ≥8 hours	119	0.97	0.74–1.26

Abbreviations: CI—confidence interval; HR—hazard ratio; HUNT—The Trøndelag Health Study; LC—lung cancer; NSCLC—non-small cell lung cancer; SCLC—small cell lung cancer.

[1]Adjusted for sex, BMI, smoking, passive smoking, leisure-time physical activity, alcohol consumption, education, economic difficulties, family history of cancer, and chronic bronchitis. Age was used as the time scale. T*he tvc* option of the *stcox* command in Stata was used to model the non-proportional hazards for smoking and economic difficulties in lung cancer overall.

Compared to participants who sat for <8 hours and were physically active, participants who sat for ≥8 hours and were physically inactive had an increased incidence of lung cancer overall and of SCLC (overall: adjusted HR =1.65, 95% CI [1.22, 2.21]; SCLC: adjusted HR = 2.58, 95% CI [1.23, 5.41]) (Figure 10). Neither of the groups with physical inactivity only or prolonged sitting only was associated with lung cancer incidence. Similar results were found among ever smokers.



SCLC—small cell lung cancer; NSCLC—non-small cell lung cancer; physically active—leisure-time physical activity level from low to high; physically inactive—reported no leisure-time activity or light activity ≤2 hours per week. Adjusted for sex, BMI, smoking, passive smoking, alcohol consumption, education, economic difficulties, family history of cancer, and chronic bronchitis. Age was used as the time scale. T*he tvc* option of the *stcox* command in Stata was used to model the non-proportional hazards for smoking and economic difficulties for lung cancer overall.

Figure 10. The association of the combined categories of total sitting time daily and leisure-time physical activity with the incidence of lung cancer overall and its major histologic types. Data from the HUNT Study, 1995–97 to 2014 (N = 33,793).

In the sensitivity analyses, we first used occupational activity as a proxy for a sedentary lifestyle and found no association between occupational inactivity (mostly sedentary work) and lung cancer incidence. Including low-level leisure-time PA in the physical inactivity group caused a weakening in the association between the combined category of sitting time of ≥8 hours and physical inactivity and the incidence of lung cancer overall and of its subtypes compared to the original results (overall: adjusted HR = 1.40, 95% CI [1.03, 1.90]; SCLC: adjusted HR = 1.99, 95% CI [0.85, 4.67]).

When excluding the first three years of follow-up, the results on the association between combined sitting for ≥8 hours and being physically inactive and the incidence of lung cancer overall and its major subtypes was unchanged. A similar trend was observed when excluding the first five years of follow-up, but statistical power was reduced.

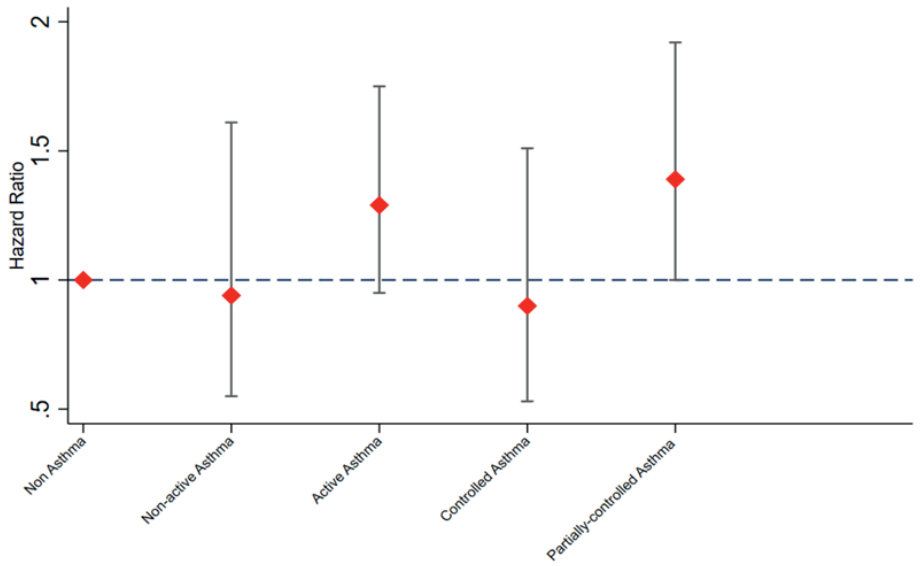
A complete case analysis for smoking found comparable estimates of the association of combined sitting for ≥8 hours and being physically inactive with the incidence of lung cancer overall. As complements to the published Paper I, multiple imputation for missing data for all covariates, including smoking, also returned comparable results to those before imputation, both for the primary cohort, and when excluding the first three years of follow-up.

Finally, when we used migraine as a negative control exposure for the association between migraine and lung cancer incidence as an additional analysis in Study1 (n = 32,821), we found that fewer participants with migraine were heavy smokers (>20.1 pack-years) compared to participants without migraine (5.8% vs 9.4%). However, no clear association between migraine and lung cancer incidence were found either before or after adjustment for smoking. Further, competing risk analysis (Fine-Gray model), another additional analysis in Study 1, showed a similar trend as our main results for the sub-cohort when excluding the influence of death, with wider CIs due to many cases of death (n = 4579).

4.2 Study 2

In total, 984 of the 62,791 participants developed lung cancer during a median follow-up time of 21.1 years. Compared to those without asthma, participants with non-active or active asthma were more likely to be former or passive smokers, and to have allergic rhinitis, economic difficulties, or a family history of cancer at baseline. Similar patterns were found among participants with controlled or partially controlled asthma compared with participants without asthma. The proportion of controlled asthma was 37% among the participants with asthma, and 29% among the participants with active asthma.

There was no clear association between asthma overall and lung cancer incidence, with an HR of 1.19 (95% CI [0.91, 1.57]) after adjustment for smoking and other confounders. Active asthma tended to be associated with an increased lung cancer incidence with an imprecise estimate (HR = 1.29, 95% CI [0.95, 1.75]). Notably, partially controlled asthma showed an increased lung cancer incidence, with an adjusted HR of 1.39 (95% CI [1.00, 1.92]). Neither non-active asthma nor controlled asthma was associated with the incidence of lung cancer (Figure 11). In the adjusted model, allergic rhinitis as a possible confounder was not associated with lung cancer incidence (HR 1.05, 95% CI [0.80, 1.39]). In addition, participants with both asthma and allergic rhinitis had a similar HR for lung cancer incidence (HR = 1.25, 95% CI [0.86, 1.82]) as participants with asthma and without allergic rhinitis (HR = 1.22, 95% CI [0.81, 1.83]) compared with those without asthma.



In the adjusted model, sex, BMI, smoking and passive smoking, alcohol consumption, leisure-time physical activity, total sitting time daily, education, economic difficulties, family history of cancer, and allergic rhinitis were included as confounders. Age was used as the time scale. T*he tvc* option of the *stcox* command in Stata was used to model the non-proportional hazards for sex, smoking, and economic difficulties.

Figure 11. The associations of asthma status and asthma symptom control with incidence of lung cancer overall. Data from the HUNT Study, 1995–97 to 2017 (N = 62,791).

From our additional analysis of the lung cancer subtypes during thesis writing (Table 3), we found a positive association between partially controlled asthma and the incidence of NSCLC (HR = 1.57, 95% CI [1.05, 2.34]). However, partially controlled asthma was not associated with SCLC (HR = 0.66, 95% CI [0.23, 1.84]).

Table 3. The associations of asthma status and asthma symptom control with the incidence of major lung cancer subtypes, the HUNT Study, 1995-97 to 2017 (N = 62,791)

				Adjusted[1]	
			Cases	HR	95% CI
SCLC	Asthma overall	No	143	1.00	Reference
		Yes	9	0.89	0.43–1.82
	Asthma status	Non-active asthma	3	1.22	0.38–3.92
		Active asthma	6	0.78	0.33–1.84
	Asthma symptom control	Controlled	5	1.73	0.68–4.38
		Partially controlled	4	0.66	0.23–1.84
NSCLC	Asthma overall	No	565	1.00	Reference
		Yes	38	1.20	0.84–1.70
	Asthma status	Non-active asthma	9	0.98	0.50–1.92
		Active asthma	29	1.29	0.87–1.91
	Asthma symptom control	Controlled	6	0.58	0.26–1.32
		Partially controlled	28	1.57	1.05–2.34

Abbreviations: CI–confidence interval; HR–hazard ratio; NSCLC–non-small cell lung cancer; SCLC–small cell lung cancer.

[1] Adjusted for sex, body mass index, smoking, passive smoking, alcohol consumption, physical activity, total sitting time daily, education, economic difficulties, family history of cancer, and allergic rhinitis. Age was used as the time scale. The *tvc* option of the *stcox* command in Stata was used to model the non-proportional hazards for smoking, economic difficulties, BMI, and allergy in the adjusted models for SCLC; *tvc* option of the *stcox* command in Stata was used to model the non-proportional hazards for sex, smoking, economic difficulties, and BMI in the adjusted models for NSCLC.

In the sensitivity analyses, when we excluded the first three years of follow-up, the associations between active asthma and partially controlled asthma and the incidence of lung cancer were stronger (HR = 1.41, 95% CI [1.04, 1.93] for active asthma, and HR = 1.54, 95% CI [1.10, 2.14] for partially controlled asthma). After excluding the first five years of follow-up, a similar pattern of results emerged, albeit with wider CIs. Likewise, similar results were

found after the exclusion of asthma cases with a higher possibility of COPD (using a post-bronchodilator fixed ratio of FEV_1/FVC or the lower limit of normal approach). Multiple imputation for missing data for all covariates including smoking also found comparable association estimates compared with those before imputation, both for the primary cohort, and when excluding the first three years of follow-up. Further, the analysis of migraine as a negative control exposure (n = 49,945) suggested that the associations we found between partially controlled asthma and the incidence of lung cancer overall were less likely to be biased by residual confounding due to smoking. Finally, the results of the competing risk analysis were similar to our main results, although they had wider CIs due to the many cases of death (n = 15,653).

4.3 Study 3

In total, 1009 of the 62,453 participants in Study 3 developed lung cancer during a median follow-up time of 21.1 years. Of these cases, 327 were lung adenocarcinoma. Compared to those with a BMI of <25 kg/m^2, participants with a BMI of 25.0–29.9 or ≥30.0 kg/m^2 were older, more likely to be former smokers and non-drinkers, less active and less educated, and more likely to have a family history of cancer at baseline.

In the observational part of Study 3, BMI in HUNT2 was inversely associated with the incidence of lung cancer overall after adjustment for smoking and other confounders. The HRs were 0.79 (95% CI [0.69, 0.91]) for a BMI of 25.0–29.9 kg/m^2 and 0.75 (95% CI [0.62, 0.91]) for a BMI of ≥30.0 kg/m^2, compared with a BMI of <25.0 kg/m^2. There was a stronger inverse association between BMI and adenocarcinoma (P for trend <0.001), with HRs of 0.73 (95% CI [0.58, 0.92]) and 0.53 (95% CI [0.37, 0.76]), for the BMIs of 25.0–29.9 kg/m^2 and ≥30.0 kg/m^2 respectively. There was no clear association between BMI in HUNT2 and the incidence of SCLC or squamous cell lung cancer. Additional adjustment for asthma, alcohol consumption, and occupational activity did not change the results markedly.

The results of the observational sensitivity analyses provided supportive evidence for the above findings. First, after excluding the first five years of follow-up, the association estimate of BMI in HUNT2 was slightly attenuated for lung cancer overall (*n* = 854) but remained similar for adenocarcinoma (*n* = 281). Second, the competing risk analysis returned similar results to our primary results, despite the many cases of death (*n* = 15,472). Third, multiple imputation for all covariates with missing data, including smoking, showed a comparable association estimate between BMI in HUNT2 and adenocarcinoma. Fourth, in the analysis using migraine as a negative control exposure (*n* = 49,969), no clear association between migraine and incidence of adenocarcinoma was identified, suggesting our observed inverse association of BMI in HUNT2 with adenocarcinoma was less likely to be biased by residual confounding from smoking.

Despite there was an inverse association between BMI measured at one time point in HUNT2 and adenocarcinoma, there was little evidence of a dose–response relationship between BMI change from HUNT1 to HUNT2 in quartiles and the incidence of adenocarcinoma (P for trend = 0.08). Compared to participants with a BMI change of -21.3–0.5 kg/m^2 in the 1st quartile, participants with a BMI change of 0.6–1.7, 1.8–3.1, and 3.2–18.6 kg/m^2 had HRs of 0.68 (95% CI [0.49, 0.92]), 0.68 (95% CI [0.49, 0.94]), and 0.78 (95% CI [0.56, 1.09]), respectively. Compared to participants who had information on BMI change from HUNT1 to HUNT2 (*n* = 44,393), those without information on the BMI change (*n* = 18,060) were younger, more physically active, and had higher socio-economic status.

We further performed MR analyses with 54,511 of the 62,453 participants with available BMI-associated SNPs. Most of the baseline characteristics were similar between participants with (*n* = 54,511) and without (*n* = 7942) information on SNPs. The exception was alcohol consumption and leisure-time PA level, for which there were relatively large differences

between the two populations. The multivariable MR estimates suggested a positive association between genetically determined BMI and the incidence of adenocarcinoma after genetically controlling for smoking. Each 1 kg/m^2 increment in genetically determined BMI directly increased the incidence of adenocarcinoma by 28% (HR = 1.28, 95% CI [1.03, 1.58]) when using the multivariable IVW method. The results using the multivariable MR-Egger method were similar (HR = 1.45, 95% CI [1.01, 2.09]). There were no clear associations between genetically determined BMI and the incidence of other lung cancer subtypes. The F-statistic value of the GRS including 75 BMI-associated SNPs was 1174 and explained 2% of the variance in BMI in HUNT2. The conditional F-statistic was 196, which suggests that the SNPs used in the multivariable MR were good instruments for BMI conditional on smoking. Results from the Cochran's Q test for adenocarcinoma suggested possible heterogeneity of the ratio estimates (P for Q = 0.01). After removing one outlier SNP (rs2121279), identified using the MR-PRESSO method, no heterogeneity between the remaining ratio estimates was observed (P for Q = 0.09). Using the intercept test from MR-Egger, pleiotropy was found to be balanced, as the P value for the intercept test was >0.05. The univariable MR analyses (as sensitivity analysis) using the 61 BMI-only SNPs returned results similar to those from the multivariable MR analyses with wider 95% CIs.

5 DISCUSSION

5.1 Summary of main findings

In this thesis, we have investigated the prospective associations between risk factors other than smoking (i.e., prolonged sitting, its combination with leisure-time PA, asthma, asthma symptom control, BMI, and BMI change) and the incidence of lung cancer overall and its histologic types. In Study 3, we also used several MR approaches to explore the potential causal association between genetically determined BMI and incidence of lung cancer overall and its subtypes. The main findings can be summarized as follows:

1) We did not find a clear association between total sitting time daily and incidence of lung cancer overall or its major subtypes. However, compared with participants sitting for <8 hours daily and being physically active, participants sitting for ≥8 hours daily and being physically inactive appeared to have increased incidences of lung cancer overall and SCLC.
2) We did not observe a clear association between asthma overall and incidence of lung cancer overall or its subtypes. However, partially controlled asthma was associated with an increased incidence of lung cancer overall and NSCLC. Adults with active asthma showed a tendency toward increased incidence of lung cancer overall. There was no clear association between non-active asthma or controlled asthma and incidence of lung cancer overall or its subtypes.
3) In the observational part of the last study, we found BMI was inversely associated with the incidence of lung adenocarcinoma in a dose–response fashion. However, the dose–response relationship was not supported by our analysis of BMI change. In the MR part of the study, we observed a positive association between genetically determined BMI and the incidence of adenocarcinoma after taking account of

smoking genetically. There were no clear associations between BMI and the incidence of other subtypes, either in the observational or MR analyses.

5.2 Comparison with previous studies

5.2.1 Study 1

In this study, we did not observe a clear association between total sitting time daily and lung cancer incidence. This is consistent with the findings of Wang et al. [28]. However, that study only included menopausal women, and it did not adjust for leisure-time PA. In contrast, we extended the findings to both men and women and adjusted for different levels of leisure-time PA.

Previous studies used different markers for a sedentary lifestyle and showed different results for the association between a sedentary lifestyle and lung cancer risk [22-26, 28]. Occupational sitting was shown to be either protective [22, 24] or not associated with lung cancer risk [23]. Possible confounding by PA or education might explain the differences in these results. On the contrary, leisure-time TV watching was suggested to be associated with an increased risk of lung cancer among Japanese men but not women [25]. Japanese women seemed to be more active than Japanese men (4.5 hours of housework for women vs 1 hour for men daily), which might explain the gender difference in lung cancer risk. Residual confounding by smoking is likely to be another explanation since there was a difference in smoking habits between Japanese men and women.

We further examined the influence of a sedentary lifestyle on lung cancer incidence using a combination of total sitting time daily and leisure-time PA level. An increased incidence of lung cancer was observed among the most sedentary individuals, who were both seated for extended periods and physically inactive. Our finding was to some extent supported by Lam et al. [26], who found a marginally increased risk of lung cancer associated with prolonged sitting among non-smokers [26]. Lam et al. largely avoided the influence of confounding due

to smoking by focusing on non-smokers. However, their results might be underestimated, and a more obvious positive association may have been seen if they had used a definition to catch the most sedentary group of people. Leisure-time TV watching, and occupational inactivity were also studied by Lam et al., but no associations with lung cancer incidence were found. In our study, neither prolonged sitting nor occupational inactivity were independently associated with lung cancer incidence after adjustment for detailed categories of smoking and leisure-time PA. This indicates that the currently suggested markers for a sedentary lifestyle may be insufficient. Future studies that define a sedentary lifestyle using objective measurement such as accelerometer might be useful. In a recent study by Parada et al., accelerometer-measured PA was inversely associated with lung cancer incidence and even replacing 30 minutes of sedentary time with light PA might reduce the lung cancer incidence [133]. However, based on a new MR study, there was little evidence that PA was associated with lung cancer incidence by using the genetical variants for PA as exposure [134]. The discrepancy between the current findings warrants further investigation in future research.

5.2.2 Study 2

In this study, we did not observe a clear association between asthma overall and incidence of lung cancer overall. This was inconsistent with previous meta-analysis studies that suggested a positive association between asthma and lung cancer [37-39]. One explanation for this discrepancy may be residual confounding by smoking. Many of the studies included in the meta-analyses did not thoroughly address the role of smoking in the association [40-42].

We further found that partially controlled asthma was associated with an increased incidence of lung cancer. Participants with active asthma seemed to have a tendency for increased lung cancer incidence in our study. In our additional analysis for subtypes in this thesis, we found partially controlled asthma was associated with an increased incidence of NSCLC but not with SCLC. In line with our findings, Pirie and his colleagues, whose study included over

half a million never smoking women from the UK Million Women Study who were followed up for more than 14 years, found that asthma requiring treatment was associated with an increased incidence of lung cancer [50]. In our study, 29% of the active asthma patients were controlled. As there was no association between controlled asthma and the incidence of lung cancer, this may have diluted the association so that we only observed a tentative association between active asthma and the incidence of lung cancer. Thus, the degree of asthma symptom control may better reflect the levels of inflammation in the airways compared to the definition of active asthma.

However, participants with partially controlled asthma were more likely to visit their physicians than the controlled asthma group. This may have contributed to an increased rate of referral for x-rays of the thorax and thus a greater chance of screening for lung cancer. In addition, 50% of the participants with partially controlled asthma in our study reported having used ICS regularly. Previous studies have reported an independently inverse association between the use of ICS and lung cancer risk [135, 136]. However, our study showed similar HRs for lung cancer incidence among adults with partially controlled asthma who used ICS compared with those who did not use ICS (data not presented). As we did not have information on the dosage or the patients' compliance with ICS use, the potential influence of ICS on the risk of lung cancer warrants further investigation.

5.2.3 Study 3

In the observational part of the study, we observed an inverse association between BMI and the incidence of lung adenocarcinoma. The inverse association did not seem to be explained by residual confounding from smoking. This finding was consistent with previous studies [61, 67, 69]. Our finding was also supported by a recent study with large number of lung cancer cases (6735) and a long follow-up duration (20 years) [67]. This study suggested that

residual confounding or within-person variability in smoking seemed unlikely to be the reason for the observed inverse association between BMI and lung cancer incidence.

Nevertheless, the observed inverse association between BMI and adenocarcinoma did not seem to be a causal association for the following reasons. First, unlike the results of BMI at baseline, our results of BMI change over 10 years in adulthood did not show a dose–response relationship with the incidence of adenocarcinoma. Second, we might not have completely excluded the residual confounding by smoking even though we had attempted to address it in several ways. This is supported by the findings from two large cohort studies of 1.2 million women in the UK [50, 65], in which no clear association between BMI and the incidence of lung cancer was found in never smokers. Third, unmeasured or unknown confounders always exist in observational studies. Obesity is accompanied by many other lifestyle factors, some of which may not be measured or remain unknown. Fourth, our multivariable MR analysis suggested a positive association between BMI and the incidence of adenocarcinoma, rather than the inverse association indicated by the observational analysis.

Among the limited MR studies, a positive association has been suggested between BMI and small or squamous cell lung cancer, while the findings on adenocarcinoma have been inconsistent [74, 75, 79, 137]. The majority of studies have applied a univariable MR approach [74, 79]. Since the interaction between BMI and smoking in the development of lung cancer seemed complicated [75, 79], the assumption of no horizontal pleiotropy might be violated in the univariable MR. It is possible to jointly examine the effects of BMI and smoking on lung cancer using a multivariable MR because it can better address horizontal pleiotropy, when smoking is either a potential confounder or mediator [75, 77]. So far, only one multivariable MR study, using a two-sample design, has been conducted. In contrast to our findings, it suggested an inverse association between BMI and the risk of adenocarcinoma after taking account for smoking [75]. However, this referred study might

suffer from weak instrument bias since its conditional F-statistic was below 10 [75]. In contrast, our conditional F-statistic was 196, indicating that the genetic variants in our study were better instruments for BMI after genetically controlling for smoking. Further, the assumption of independence could not be checked in a two-sample MR study. Conversely, our study was a one-sample MR, which allowed a thorough check for associations between BMI-associated SNPs and the important confounders. Therefore, we were also able to further adjust for COPD when calculating the association estimates between BMI-only SNPs and adenocarcinoma. The causal estimates from our multivariable MR using all the BMI-associated SNPs and the univariable MR using the BMI-only SNPs showed similar trends.

Taken together, our findings suggest that the observed inverse association between BMI and lung adenocarcinoma may not be causal. Residual confounding by smoking was less likely to be the reason for the observed inverse association between BMI and lung adenocarcinoma. More MR studies are needed to confirm our finding of a positive association between BMI and lung adenocarcinoma.

5.3 Methodological considerations

Precision and validity are two important components of accuracy in epidemiological studies [138]. When interpreting our results, we needed to consider both random error and systematic error since random error might reduce the precision of the association (by causing chance findings) and systematic error could interfere with the validity of the results (via bias). Validity can be divided into internal and external validity. Internal validity is a prerequisite for external validity [138]. There are several types of bias in observational studies that threaten internal validity, such as selection bias, information bias, and confounding. External validity is how generalizable the study is to other populations [138]. More detailed discussion of these methodological considerations follows.

5.3.1 Random error and statistical precision

Random error refers to the chance difference between observed values and the true value [138]. The 95% CI is used to measure random error. A wider CI indicates lower statistical precision and a higher possibility of random error. Random error can be reduced by increasing the sample size in epidemiological studies. In this thesis, we had relatively large study samples in all three studies, which ensured our main findings were relatively precisely estimated, as indicated by narrower CIs. However, in Study 1, the sample size of the main cohort (n = 45,810) and especially of the sub-cohort (n = 33,793) was smaller and the follow-up duration was shorter than that of the Studies 2 and 3. This was unfavorable for analysis for lung cancer subtypes. It was thus decided to perform the analysis for the main subtypes, SCLC and NSCLC.

Compared with Study 1, Studies 2 and 3 had larger sample sizes (62,791 and 62,453 participants, respectively) and a longer follow-up time. This increased the precision of the association estimates in the main analyses and also in the analysis between migraine and lung cancer incidence (for the negative control exposure analysis) in these studies compared to Study 1. Although we only presented results for lung cancer overall in Paper II, the sample size was large enough for us to do additional analysis for SCLC and NSCLC separately for the thesis. In Study 3, due to the large sample size and a possible stronger association between BMI and lung cancer, we were able to investigate the associations for both lung cancer overall and its specific subtypes. We were also able to present clear estimates of associations when excluding the first five years of follow-up for investigating potential reverse causation. On the contrary, the estimates of associations were not clear when we excluded the first five years of follow-up in Studies 1 and 2. Finally, although we could not investigate the associations among never smokers in any of the three studies due to fewer lung cancer cases (lung cancer cases among never smokers were 26 in Study 1, 56 in Study 2,

and 55 in Study 3), we were able to address residual confounding by smoking using a range of methods (as discussed further in Section 6.2.4).

5.3.2 Selection bias

Selection bias refers to the differences in exposure–outcome associations between the people participating in a study and those eligible for participation [138]. In this thesis, selection bias due to non-participation, missing data, and competing risk of death was evaluated.

Non-participation: In HUNT2, all adults (20 years or older) in the north area of Trøndelag were invited to the study, and around 70% of those eligible participated. This is a relatively high participation rate, which reduces selection bias. In addition, a non-participation study was performed, which found that the main reasons for non-participation were lack of time and moving to other places [94]. In general, no major systematic difference was apparent between participants and non-participants in HUNT2. But participants tended to be healthier than non-participants, especially for the older participants [93].

Missing data: In Study 1, selection bias cannot be completely ruled out. Among the cancer-free population (n = 59,070), 22% (13,260/59,070) of participants were excluded due to missing information on total sitting time daily. Compared to the cancer-free population, participants included in the main cohort (n = 45,810) had a higher percentage of family history of cancer which was taken into adjustment in the statistical models. Further, 26% (12,017/45,810) of participants were excluded in the sub-cohort due to a lack of information on leisure-time PA. Nevertheless, there were no major differences between the included and excluded participants in the sub-cohort which suggested no substantial selection bias. In Study 2 and 3, the main exposures such as asthma and BMI had very few missing, thus selection bias might not have a substantial impact to our main results in these two studies. In Study 3, 13% (7942/62,453) of participants were excluded in the sub-cohort for MR analyses and the excluded participants were shown to have lower alcohol consumption and lower

leisure-time PA levels than the included participants. However, the proportions of those with a BMI of ≥25.0 kg/m^2 (59.9 % vs 60.0 %) and of lung cancer cases (1.6% vs 1.7%) were similar between the included and excluded participants. Thus, selection bias might not be a concern for the observed associations in the sub-cohorts.

Missing data on covariates is normal in prospective studies, but may introduce selection bias [139]. As many people (8.2%) were missing data on smoking, the most important risk factor for lung cancer, we undertook a complete case analysis for smoking in Study 1 to address the residual confounding by smoking. However, this method has some limitations. First, smaller sample sizes may reduce statistical power and precision. Second, missing covariate data may be missing at random, or missing not at random. If we assume missing is completely at random, missing data is less of a problem, as there would be no systematic difference between what has been missing and what has been observed. If the missing is not at random, complete case analysis may systematically bias our results. In this thesis, data was also largely missing on leisure-time PA (26.2% in Study 1, 30.4% in Study 2 and 30.4% in Study 3) and economic difficulties (13.5% in Study 1, 30.3% in Study 2 and 29.9% in Study 3). All missing values of the covariates were included in the analysis as an "unknown" category. This approach would be appropriate if missing was not at random and determined by the value itself. We further performed multiple imputation for all covariates with missing data based on the assumption that the data was missing at random. The results after imputation demonstrate there was no serious bias due to missing data despite the large proportion of missing for some parts of the study.

Competing risk of death: Loss to follow-up is another source of selection bias. However, except for competing risk of death during follow-up, this type of selection bias is unlikely to have been a problem in this thesis, as the migration rate in HUNT2 was less than 0.3% per

year and all lung cancer cases are tracked via the unique personal identification number linked to the Cancer Registry of Norway which is a national register [9, 94].

As mentioned in the statistical analysis section (see Section 4.4), competing risk of death during follow-up may cause biased association estimates. As many people diagnosed with lung cancer are already in the advanced stage of the disease, other serious comorbidities may exist at the same time. Thus, people may die of other reasons rather than lung cancer during the follow-up duration. To deal with possible bias due to competing risk of death, we performed competing risk analyses for death in all three studies, with the results being comparable to our main results.

5.3.3 Information bias

Information bias happens when subjects are incorrectly categorized, their exposure or outcome status is misclassified, thereby potentially altering the observed association estimate away from the true value[138]. This is also called misclassification, and it may be differential or non-differential [138]. Differential misclassification happens when the misclassification of the exposure or the outcome is related to the outcome or the exposure, respectively. Bias from differential misclassification can either overestimate or underestimate the true association [138]. When misclassification occurs equally across two groups: exposed and unexposed or diseased and non-diseased groups in the study, the misclassification is non-differential and generally attenuates of the true association [138].

Misclassification of exposures: Exposures such as total sitting time daily and leisure-time PA in Study 1, and asthma and asthma status in Study 2, were self-reported, giving rise to the risk of misclassification. However, the categorization of total sitting time used in this thesis is comparable with other HUNT studies [96, 97] and two meta-analysis studies [98, 99]. The measurement of leisure-time PA in HUNT2 was also validated [140]. Self-reported asthma has been verified to be highly specific and reliable in many observational studies [141]. In

addition, the prevalence of asthma and active asthma (5.1% and 3.3%, respectively) in Study 2 was in line with the prevalence reported in a previous HUNT study using a slightly different definition of asthma [142] and in another Nordic study [143]. In Study 2, 37% of the participants with asthma had symptom controlled, which is comparable to the findings from other European studies [52, 53]. Notably, COPD may be misdiagnosed as asthma due to similar symptoms, especially among elderly people. This can bias the association between asthma and lung cancer. To deal with this, we excluded participants with possible COPD according to the GOLD definition at baseline and further excluded asthma cases with a higher possibility of COPD (people smoked ten pack-years or more and were older than 40 years when getting the asthma diagnosis and had a) post-bronchodilator FEV_1/FVC <0.7 based on fixed ratio criterion or b) post-bronchodilator FEV_1/FVC z score <-1.64 based on lower limit of normal criterion) in the sensitivity analyses in Study 2. The results from the sensitivity analyses were similar to our primary results, indicating that possible misclassification of COPD as asthma was unlikely to explain the observed association. In Study 3, BMI was measured objectively, and misclassification of BMI was a minor issue compared to the self-reported exposures in Studies 1 and 2. Further, all information on exposures was collected at baseline long before the diagnosis of lung cancer. As such, any misclassification of the exposures is likely to be non-differential and thus underestimate associations.

All the exposures considered in this thesis fluctuate over time. The association estimates generated from HUNT2's single time point measurement may not be a true reflection of the effect of the varying exposures on lung cancer incidence. To deal with this, we performed the analyses in Studies 1 and 2 again, treating all exposures as time varying since we also had similar exposure data in HUNT3. The results were comparable to those from the original analysis (data not shown). Further, in Study 3, genetically determined BMI was used to reflect average BMI across the lifespan.

Misclassification of outcome: All information on lung cancer and histologic types came from the Cancer Registry of Norway, which records relatively complete and accurate information about lung cancer diagnosis and different histologic types from one year after first diagnosis [106]. The quality of these data means there is a low risk of outcome misclassification in this thesis. However, we did need to consider the fact that the participants with partially controlled asthma in Study 2 were more likely to visit their physicians than participants with controlled asthma or without asthma. This may have increased their chance of screening for lung cancer, producing differential misclassification and potentially upwardly biasing the association estimates.

5.3.4 Confounding

Confounding is a major source of bias in observational studies [138]. In this thesis, we had information on a panel of potential confounders at baseline, including age, gender, different lifestyle factors, and comorbidities, which made it possible to control for bias by most confounders.

Smoking, which is the most important risk factor for lung cancer, there were indications that smoking status was distributed differentially across the exposed and unexposed groups in all three studies of the thesis. We could not stratify our analyses by smoking due to the limited number of lung cancer cases found in never smokers (26 in Study 1, 56 in Study 2 and 55 in Study 3). Therefore, to minimize the bias due to residual confounding by smoking in the thesis, we used detailed information on smoking status (never, former, and current) and pack-years to categorize the smoking variable and performed several sensitivity analyses. First, we conducted a complete case analysis for smoking in Study 1. Second, and also in Study 1, we redefined smoking to include cessation years for former daily smokers and categorized subjects into the groups of never smokers, ex-smokers with smoking cessation of >10.1 years, ex-smokers with smoking cessation of ≤10.0 years, current smokers with 0.0–20.0

pack-years and current smokers with >20.1 pack-years. Third, a negative control exposure (migraine) was used for all the studies. The results for migraine on lung cancer incidence overall are shown in Table 4. Forth, we performed multiple imputation analysis for all covariates with missing including smoking. The findings suggested that our primary results were less likely to be biased by residual confounding by smoking, especially for studies 2 and 3. For study 1, the negative control exposure result was not clear. Thus, residual confounding by smoking cannot be excluded in study 1.

Table 4. Negative control using migraine as an alternative exposure to address residual confounding by smoking in all three studies.

		Crude[1]		Adjusted[2]		Adjusted[3]	
	Cases	HR	95% CI	HR	95% CI	HR	95% CI
Study 1 (N = 32,821)							
No migraine	316	1.00	Reference	1.00	Reference	1.00	Reference
Migraine	17	0.72	0.44–1.17	0.84	0.51–1.37	0.84	0.51– 1.38
Study 2 (N = 49,945)							
No migraine	721	1.00	Reference	1.00	Reference	1.00	Reference
Migraine	35	0.61	0.43–0.86	0.75	0.53–1.06	0.75	0.53–1.05
Study 3 (N = 49,969)							
No migraine	738	1.00	Reference	1.00	Reference	1.00	Reference
Migraine	40	0.68	0.49–0.93	0.83	0.60–1.14	0.80	0.58–1.11

Abbreviations: CI—confidence interval; HR—hazard ratio.

[1] Age was used as the time scale in the crude model. [2] Adjusted for smoking (never, former [≤10.0, 10.1–20.0 and >20.1 pack-years (pyrs)], and current [≤10.0, 10.1–20.0, and >20.1 pyrs]) in all three studies. Age was used as the time scale. [3] In Study 1, we adjusted for sex, BMI, smoking, passive smoking, alcohol consumption, education, economic difficulties, family history of cancer, and chronic bronchitis. In Study 2, we adjusted for sex, BMI, smoking, passive smoking, alcohol consumption, leisure-time physical activity, total sitting time daily, education, economic difficulties, family history of cancer, and allergic rhinitis. In Study 3, we adjusted for sex, smoking, passive smoking, leisure-time physical activity, total sitting time daily, education, economic difficulties, family history of cancer, and self-reported COPD. Age was used as the time scale. The *tvc* option of the *stcox* command in Stata was used to model the non-proportional hazards in the adjusted[3] model. The non-proportional hazards for Study 1 were smoking, economic difficulties, and alcohol; for Study 2 and Study 3, they were sex, smoking, and economic difficulties.

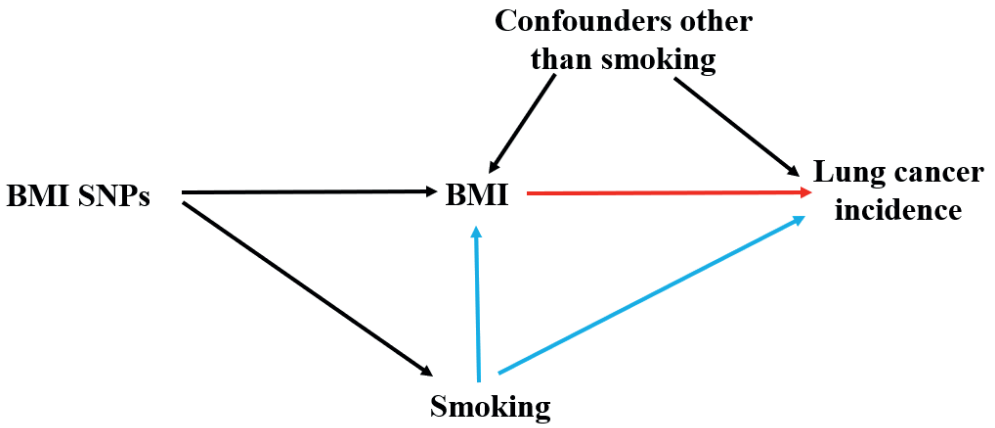
In the negative control exposure analysis, we found that combining smoking status with detailed information on pack-years allowed for greater precision than if using the simpler smoking status categories of never, former, and current. For instance, in Study 2, the rate of migraines was similarly distributed across the different smoking categories (8.6% among never smokers, 7.8% among former smokers and 8.8% among current smokers). However, a clearer difference emerged when smoking was categorized by pack-year (8.6% of never smokers reported migraines compared to 4.3% of former smokers with >20.1 pack-years). In other words, bias due to confounding by smoking was likely reduced when we combined smoking status with detailed information on pack-years to create more detailed categories.

Still, many confounders may have measurement errors or missing data, and residual confounding can be introduced and distort the association between exposures and outcomes even after adjusting for confounders. Further, there are always unmeasured confounders. For example, we have no information on air pollution, indoor radon, or occupational exposure to asbestos or other carcinogenic agents, which have all been identified as important risk factors for lung cancer [144]. However, we can assume that the risk from these potential confounders is minimal since, except for the two smallest municipalities, which have had mining industries previously, there was nearly no industrial pollution in the northern area of Trøndelag during the time of the HUNT2 Study [142]. Likewise, when national measurements were taken between 1999–2000, the level of indoor radon in the north area of Trøndelag was found to be in the safe range (<200 Bq/m^3) [145]. Finally, no asbestos-cement factories have ever existed in the county [146], and the country-wide prohibition on the importation of asbestos and strict regulation of its use since 1980 means that vanishingly few people in the HUNT Study would have had heavy exposure to asbestos [146].

Even though we performed several sensitivity analyses to address residual confounding by smoking in our thesis, it might not have been excluded entirely due to the nature of the

observational study design. Some unmeasured or unknown confounders, may or may not be associated with the measured confounders, also threaten the validity of the epidemiologic study [147].

However, the MR approach helps to control for residual confounding or unmeasured confounding. In Study 3, we performed a multivariable MR in which 75 genetic variants for BMI (including 14 genetic variants also related to smoking) were used as instruments for BMI and smoking (Figure 12). Given the MR assumptions hold, this study likely avoided residual and unmeasured confounding by smoking, and a positive causal association for BMI on lung cancer was found. The discrepancy in our results between the observational and MR analyses in Study 3 likely demonstrates the problematic nature of residual confounding or unmeasured confounding in observational studies.



BMI—body mass index; MR—Mendelian randomization; SNPs—single nucleotide polymorphisms.

Since 14 of the 75 BMI SNPs are also suggested to be associated with smoking with a strict P value of <0.05, the influence of smoking can be controlled by using all 75 BMI-associated SNPs as instruments for BMI in the multivariable MR analysis. BMI and smoking are both regarded as exposures in this situation.

Figure 12. Association between BMI and the incidence of lung cancer, genetically controlling for the influence of smoking using multivariable MR.

5.3.5 Reverse causation

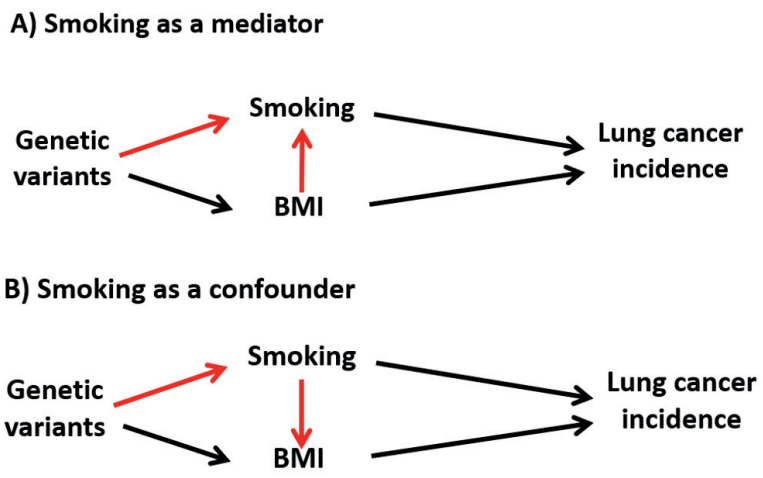
Reverse causation refers to a situation in which the outcome or proximal precursor of the outcome, such as pre-clinical disease, causes the exposure, rather than the exposure causing the outcome [148]. Since people who are diagnosed with lung cancer are always at a late stage of the disease, the latent period of lung cancer is usually long. During the latent period, people may already be experiencing the symptoms of lung cancer, such as tiredness, lack of energy, and cough, which in turn causes a more sedentary lifestyle or reduces body weight. More specifically, it is difficult to distinguish whether people have asthma or lung cancer since some of the symptoms are similar. Thus, reverse causality may exist in all three studies in the thesis, leading to biased estimates of associations. To deal with this, we excluded the first three years of follow-up in Studies 1 and 2, and the first five years of follow-up in Study 3. The results indicated that potential reverse causation might not explain the observed associations. Further, using an MR approach, we were able to avoid bias due to reverse causation, since genetic variants are fixed from conception. The results from both our univariable and multivariable MR analyses suggested that positive association between genetically determined BMI and lung adenocarcinoma in Study 3 was unlikely due to reverse causation.

5.3.6 Pleiotropy

Pleiotropy is a common phenomenon in the human genome in which a single genetic variant influences multiple traits [149]. When the genetic variant influences one trait and in turn influences another, it is called vertical pleiotropy [149]. In this case, the exclusion assumption of the MR study would not be violated because the genetic variant influences the outcome only through the exposure and no bias exists in the causal estimate. Conversely, in horizontal pleiotropy, which is when the genetic variant influences the outcome through an independent pathway other than the exposure, the causal estimate in the MR study could be biased [149].

When multiple SNPs are used, it is possible to perform the MR-Egger and MR-PRESSO methods among others, to test for horizontal pleiotropy by modeling or removing outliers if any exist respectively. If the horizontal pleiotropic effects of SNPs are equivalently positive and negative, this suggests that the pleiotropy is balanced and the MR estimate is unbiased [149].

In Study 3, there may be two scenarios between BMI and smoking in relation to lung cancer (Figure 13): A) smoking as a mediator in the BMI-lung cancer association; B) smoking as a confounder in the BMI-lung cancer association. We assumed scenario B was more likely based on previous research [69]. Horizontal pleiotropy may be present in both scenarios since GWAS suggested common genetic variants for BMI and smoking [75, 76]. In our Study 3, we found that 14 BMI-associated SNPs were also associated with smoking. Thus, the causal estimate between BMI and lung cancer could be biased if univariable MR was used since it cannot handle horizonal pleiotropy.



BMI—body mass index.

A: Smoking as a mediator; B: Smoking as a confounder. The pleiotropic associations are shown by the red line.

Figure 13. Directed acyclic graphs showing possible horizontal pleiotropy caused by smoking when genetic variants for BMI are also associated with smoking.

Multivariable MR, which is analogous to a factorial randomized trial that estimates the causal effects of each risk factor simultaneously [78], can be applied to handle the horizontal pleiotropy and provide unbiased causal estimates. Thus, we used multivariable MR to control for the possible horizontal pleiotropy by smoking in Study 3 (Figure 13). However, we could not exclude the possibility of horizontal pleiotropy due to unmeasured confounders since there was evidence of heterogeneity in the ratio estimates in both the multivariable and univariable MR analyses before an outlier was removed. Nevertheless, the intercept from the multivariable MR-Egger, which is sensitive to strong horizontal pleiotropy, did not indicate strong or unbalanced pleiotropic effects. After removing the outlier SNP, no heterogeneity was shown, and the MR estimates for the BMI–lung adenocarcinoma association were similar to the original results.

5.3.7 Causality in epidemiological studies

Epidemiological studies aim to clarify causality; however, it is challenging to obtain reliable causal evidence for risk factors other than smoking due to bias caused by residual confounding or reverse causation. In addition, many studies suffer from limited statistical power in subgroups such as never smokers or lung cancer cases of a specific subtype. Although prolonged sitting combined with physically inactive and partially controlled asthma seemed to be associated with an increased incidence of lung cancer in the first two studies, we could not draw a conclusion on causality.

Well-conducted RCTs are regarded as the gold standard for causal evidence of treatment [138]. However, it is not always feasible or ethical to perform an RCT, especially when studying the causal effect of an exposure on cancer. Further, the findings from some large RCTs contradict expectations based on observational data. For example, the Selenium and Vitamin E Cancer Prevention Trial found no effect of selenium, vitamin E, or a combination of both on preventing prostate cancer [150]. This highlights the importance of obtaining valid

causal evidence and understanding the limitations of observational epidemiology when interpreting the evidence from associations between risk factors and disease before carrying out RCTs.

The MR approach offers many benefits, as it attempts to mimic an RCT using observational data, and it can address some of the important limitations in traditional observational studies by using genetic variants as a proxy for potential modifiable risk factors. However, MR studies also have their weaknesses, such as pleiotropy or weak instrument bias. It is therefore important to address any epidemiological question by comparing the results of a range of study designs with different and independent sources of bias. In Study 3, we tried to explore the association between BMI and lung cancer incidence using both a prospective cohort and MR design. The discrepancy in the results for the association between BMI and lung adenocarcinoma suggested that limitations such as residual confounding or reverse causation may have led to erroneous conclusions in previous observational studies. This emphasizes the difficulty in understanding the relationship between BMI and lung cancer and highlights the need for more MR studies or alternative causal methods to confirm our finding of a positive causal association between BMI and lung adenocarcinoma.

5.3.8 External validity

Generalization of the findings to other populations is important for external validity in epidemiological studies. Differences in demographic, ethnic, or socioeconomic characteristics may lead to problems with external validity [151]. Our research used data from a mostly homogenous Caucasian population living in the middle of Norway, which makes our study sample representative of the Norwegian population. However, the findings may be less generalizable to other ethnic groups outside Scandinavia.

6 CONCLUSION AND FUTURE PERSPECTIVES

In this thesis, we provided some new evidence of potential risk factors other than smoking for lung cancer incidence. We observed a positive association between prolonged sitting in combination with physical inactivity and the incidence of lung cancer and SCLC. Intervention targets could therefore focus on the most sedentary group of people, whose risk of lung cancer may be reduced by replacing a sedentary lifestyle. Our study also showed that participants with only partially controlled asthma had an increased incidence of lung cancer and NSCLC. The findings suggested that proper control of asthma symptoms not only reduced asthma exacerbations but might also contribute to a reduced incidence of lung cancer. Since NCSLC is the major subtype of lung cancer and is relatively insensitive to chemotherapy, this finding has important implications for healthcare with regard to both asthma and lung cancer.

These results, however, should be interpreted with caution since residual confounding by smoking and unmeasured confounding cannot be completely excluded in observational studies. Within the framework of multivariable MR, residual confounding by smoking could be well controlled if assumptions hold, and we likely observed the direct causal association between BMI and lung cancer incidence. Unlike the inverse association between BMI and lung adenocarcinoma found in the observational analyses, our results from the multivariable MR suggested a positive association. If this positive association between BMI and the incidence of lung adenocarcinoma is proven by future MR studies, reducing body weight will come to be regarded as an important prevention target for lung adenocarcinoma.

With the increasing availability of large-scale genome-wide data from large cohort studies (e.g., UK Biobank and large genome-wide association studies for different modifiable risk factors and individual cancers), there are more opportunities to use human genetics to

disentangle causal associations between risk factors other than smoking and incidence of lung cancer overall and its subtypes. The MR approach, and the new knowledge it is generating will be a cost-effective way of improving prevention strategies for individuals at high risk of developing disease in the future.

7 REFERENCES

1. National cancer institute US. Definition for small cell lung cancer. 2006.
2. National cancer institute US. Definition for non-small cell lung cancer. 2006.
3. Syrigos KN, Nutting CM, Roussos C. Tumors of the Chest: Biology, Diagnosis and Management: Springer Science & Business Media; 2006.
4. World Health Organization. Globocan 2020-lung cancer fact sheet. Geneva: World Health Organization; 2020.
5. Thandra KC, Barsouk A, Saginala K, Aluru JS, Barsouk A. Epidemiology of lung cancer. Contemp Oncol (Pozn). 2021;25(1):45-52.
6. Howlader N, Noone A, Krapcho Me, Miller D, Brest A, Yu M, et al. SEER cancer statistics review, 1975–2016, National Cancer Institute. 2019.
7. Barta JA, Powell CA, Wisnivesky JP. Global epidemiology of lung cancer. Ann Glob Health. 2019;85(1).
8. National cancer institute US. New treatments spur sharp reduction in lung Cancer mortality Rate. 2020.
9. Cancer Registry of Norway. Cancer in Norway 2019-Cancer incidence, mortality, survival and prevalence in Norway. Oslo: Cancer Registry of Norway; 2020.
10. Tyczynski JE, Bray F, Parkin DM. Lung cancer in Europe in 2000: epidemiology, prevention, and early detection. Lancet Oncol. 2003;4(1):45-55.
11. O'Keeffe LM, Taylor G, Huxley RR, Mitchell P, Woodward M, Peters SAE. Smoking as a risk factor for lung cancer in women and men: a systematic review and meta-analysis. BMJ Open. 2018;8(10):e021611.
12. Powell HA, Iyen-Omofoman B, Hubbard RB, Baldwin DR, Tata LJ. The association between smoking quantity and lung cancer in men and women. Chest. 2013;143(1):123-9.
13. Hecht SS. Tobacco smoke carcinogens and lung cancer. JNatl Cancer Inst. 1999;91(14):1194-210.
14. Barnes J, Behrens TK, Benden ME, Biddle S, Bond D, Brassard P, et al. Letter to the Editor: Standardized use of the terms" sedentary" and" sedentary behaviours". Applied Physiology Nutrition and Metabolism-Physiologie Appliquee Nutrition Et Metabolisme. 2012;37(3):540-2.
15. Park JH, Moon JH, Kim HJ, Kong MH, Oh YH. Sedentary lifestyle: Overview of updated evidence of potential health risks. Korean J Fam Med. 2020;41(6):365.
16. Brenner DR, Yannitsos DH, Farris MS, Johansson M, Friedenreich CM. Leisure-time physical activity and lung cancer risk: A systematic review and meta-analysis. Lung Cancer. 2016;95:17-27.
17. World Health Organization. Physical inactivity: a global public health problem. 2020. Grneva; 2020.
18. Patterson R, McNamara E, Tainio M, de Sá TH, Smith AD, Sharp SJ, et al. Sedentary behaviour and risk of all-cause, cardiovascular and cancer mortality, and incident type 2 diabetes: a systematic review and dose response meta-analysis. Eur J Epidemiol. 2018;33(9):811-29.
19. Tremblay MS, Colley RC, Saunders TJ, Healy GN, Owen N. Physiological and health implications of a sedentary lifestyle. Appl Physiol Nutr Metab. 2010;35(6):725-40.
20. Thorp AA, Owen N, Neuhaus M, Dunstan DW. Sedentary behaviors and subsequent health outcomes in adults: a systematic review of longitudinal studies, 1996–2011. Am J Prev Med. 2011;41(2):207-15.

21. Shen D, Mao W, Liu T, Lin Q, Lu X, Wang Q, et al. Sedentary Behavior and Incident Cancer: A Meta-Analysis of Prospective Studies. PLoS One. 2014;9(8):e105709.
22. 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):2389-97.
23. Thune I, Lund E. The influence of physical activity on lung-cancer risk: A prospective study of 81,516 men and women. Int J Cancer. 1997;70(1):57-62.
24. Bak H, Christensen J, Thomsen BL, Tjønneland A, Overvad K, Loft S, et al. Physical activity and risk for lung cancer in a Danish cohort. Int J Cancer. 2005;116(3):439-44.
25. Ukawa S, Tamakoshi A, Wakai K, Noda H, Ando M, Iso H. Prospective cohort study on television viewing time and incidence of lung cancer: findings from the Japan Collaborative Cohort Study. Cancer Causes Control. 2013;24(8):1547-53.
26. Lam TK, Moore SC, Brinton LA, Smith L, Hollenbeck AR, Gierach GL, et al. Anthropometric measures and physical activity and the risk of lung cancer in never-smokers: a prospective cohort study. PLoS One. 2013;8(8):e70672.
27. Katzmarzyk PT, Church TS, Craig CL, Bouchard C. Sitting time and mortality from all causes, cardiovascular disease, and cancer. Med Sci Sports Exerc. 2009;41(5):998-1005.
28. Wang A, Qin F, Hedlin H, Desai M, Chlebowski R, Gomez S, et al. Physical activity and sedentary behavior in relation to lung cancer incidence and mortality in older women: The Women's Health Initiative. Int J Cancer. 2016;139(10):2178-92.
29. Schmid D, Ricci C, Behrens G, Leitzmann MF. Does smoking influence the physical activity and lung cancer relation? A systematic review and meta-analysis. Eur J Epidemiol. 2016;31(12):1173-90.
30. Zhong S, Ma T, Chen L, Chen W, Lv M, Zhang X, et al. Physical Activity and Risk of Lung Cancer: A Meta-analysis. Clin J Sport Med. 2016;26(3):173-81.
31. Marks G, Pearce N, Strachan D, Asher I, Ellwood P. The burden of asthma. The Global Asthma Report. 2018:14-37.
32. 2019 GINA Report. Global Strategy for Asthma Management and Prevention 82019 update. 2019.
33. Brenner DR, Boffetta P, Duell EJ, Bickeböller H, Rosenberger A, McCormack V, et al. Previous lung diseases and lung cancer risk: a pooled analysis from the International Lung Cancer Consortium. 2012;176(7):573-85.
34. Mart M, Peebles Jr R. Asthma-chronic obstructive pulmonary disease overlap syndrome. Curr Opin Immunol. 2020;66:161-6.
35. Langhammer A, Johannessen A, Holmen TL, Melbye H, Stanojevic S, Lund MB, et al. Global Lung Function Initiative 2012 reference equations for spirometry in the Norwegian population. Eur Respir J. 2016;48(6):1602-11.
36. Pauwels RA, Buist AS, Calverley PM, Jenkins CR, Hurd SS. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary. Am J Respir Crit Care Med. 2001;163(5):1256-76.
37. Rosenberger A, Bickeboller H, McCormack V, Brenner DR, Duell EJ, Tjonneland A, et al. Asthma and lung cancer risk: a systematic investigation by the International Lung Cancer Consortium. Carcinogenesis. 2012;33(3):587-97.
38. Santillan AA, Camargo CA, Jr., Colditz GA. A meta-analysis of asthma and risk of lung cancer (United States). Cancer Causes Control. 2003;14(4):327-34.
39. Qu YL, Liu J, Zhang LX, Wu CM, Chu AJ, Wen BL, et al. Asthma and the risk of lung cancer: a meta-analysis. Oncotarget. 2017;8(7):11614-20.
40. Boffetta P, Ye W, Boman G, Nyren. Lung cancer risk in a population-based cohort of patients hospitalized for asthma in Sweden. Eur Respir J. 2002;19(1):127-33.

41. Hwang CY, Chen YJ, Lin MW, Chen TJ, Chu SY, Chen CC, et al. Cancer risk in patients with allergic rhinitis, asthma and atopic dermatitis: a nationwide cohort study in Taiwan. Int J Cancer. 2012;130(5):1160-7.
42. Ji J, Shu X, Li X, Sundquist K, Sundquist J, Hemminki K. Cancer risk in hospitalised asthma patients. Br J Cancer. 2009;100(5):829-33.
43. Vandentorren S, Baldi I, Annesi Maesano I, Charpin D, Neukirch F, Filleul L, et al. Long-term mortality among adults with or without asthma in the PAARC study. Eur Respir J. 2003;21(3):462-7.
44. Turner MC, Chen Y, Krewski D, Ghadirian P, Thun MJ, Calle EE. Cancer mortality among US men and women with asthma and hay fever. Am J Epidemiol. 2005;162(3):212-21.
45. Brown DW, Young KE, Anda RF, Felitti VJ, Giles WH. Re: asthma and the risk of lung cancer. findings from the Adverse Childhood Experiences (ACE). Cancer Causes Control. 2006;17(3):349-50.
46. Colak Y, Afzal S, Nordestgaard BG, Lange P. Characteristics and Prognosis of Never-Smokers and Smokers with Asthma in the Copenhagen General Population Study. A Prospective Cohort Study. Am J Respir Crit Care Med. 2015;192(2):172-81.
47. Mills PK, Beeson WL, Fraser GE, Phillips RL. Allergy and cancer: organ site-specific results from the Adventist Health Study. Am J Epidemiol. 1992;136(3):287-95.
48. Talbot-Smith A, Fritschi L, Divitini ML, Mallon DF, Knuiman MW. Allergy, atopy, and cancer: a prospective study of the 1981 Busselton cohort. Am J Epidemiol. 2003;157(7):606-12.
49. Polosa R, Thomson NC. Smoking and asthma: dangerous liaisons. Eur Respir J. 2013;41(3):716-26.
50. 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.
51. Levy ML. GINA asthma strategy: what's new for 2020.
52. Cazzoletti L, Marcon A, Janson C, Corsico A, Jarvis D, Pin I, et al. Asthma control in Europe: a real-world evaluation based on an international population-based study. J Allergy Clin Immunol. 2007;120(6):1360-7.
53. Demoly P, Paggiaro P, Plaza V, Bolge SC, Kannan H, Sohier B, et al. Prevalence of asthma control among adults in France, Germany, Italy, Spain and the UK. Eur Respir Rev. 2009;18(112):105-12.
54. Castro-Rodriguez JA, Rodrigo GJ. A systematic review of long-acting β2-agonists versus higher doses of inhaled corticosteroids in asthma. Pediatrics. 2012;130(3):e650-e7.
55. World Health Organization. Obesity: preventing and managing the global epidemic. 2000. Report No.: 9241208945.
56. Chooi YC, Ding C, Magkos F. The epidemiology of obesity. Metabolism. 2019;92:6-10.
57. World Health Organization. Obesity and overweight 2020. Geneva: WHO; 2020.
58. McCafferty BJ, Hill JO, Gunn AJ. Obesity: Scope, Lifestyle Interventions, and Medical Management. Tech Vasc Interv Radiol. 2020;23(1).
59. Calle EE, Kaaks R. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. Nat Rev Cancer. 2004;4(8):579-91.
60. Abdel-Rahman O. Pre-diagnostic Body Mass Index Trajectory in Relationship to Lung Cancer Incidence and Mortality; Findings From the PLCO Trial. Expert Rev Respir Med. 2019;13(10):1029-35.
61. Yu D, Zheng W, Johansson M, Lan Q, Park Y, White E, et al. Overall and Central Obesity and Risk of Lung Cancer: A Pooled Analysis. J Natl Cancer Inst. 2018;110(8):831-42.

62. Yang Y, Dong J, Sun K, Zhao L, Zhao F, Wang L, et al. Obesity and incidence of lung cancer: a meta-analysis. Int J Cancer. 2013;132(5):1162-9.
63. Tarnaud C, Guida F, Papadopoulos A, Cénée S, Cyr D, Schmaus A, et al. Body mass index and lung cancer risk: results from the ICARE study, a large, population-based case–control study. Cancer Causes Control. 2012;23(7):1113-26.
64. Zhu H, Zhang S. Body mass index and lung cancer risk in never smokers: a meta-analysis. BMC Cancer. 2018;18(1):635.
65. Reeves GK, Pirie K, Beral V, Green J, Spencer E, Bull D. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. BMJ. 2007;335(7630):1134.
66. Taylor AE, Richmond RC, Palviainen T, Loukola A, Wootton RE, Kaprio J, et al. The effect of body mass index on smoking behaviour and nicotine metabolism: a Mendelian randomization study. Hum Mol Genet. 2018;28(8):1322-30.
67. Wood AM, Jonsson H, Nagel G, Häggström C, Manjer J, Ulmer H, et al. The Inverse Association of Body Mass Index with Lung Cancer: Exploring Residual Confounding, Metabolic Aberrations and Within-Person Variability in Smoking. Cancer Epidemiol Biomarkers Prev. 2021;30(8):1489-97.
68. Chiolero A, Faeh D, Paccaud F, Cornuz J. Consequences of smoking for body weight, body fat distribution, and insulin resistance. Am J Clin Nutr. 2008;87(4):801-9.
69. Yang Y, Dong JY, Sun KK, Zhao L, Zhao F, Wang LL, et al. Obesity and incidence of lung cancer: A meta-analysis. Int J Cancer. 2013;132(5):1162-9.
70. Kanashiki M, Sairenchi T, Saito Y, Ishikawa H, Satoh H, Sekizawa K. Body mass index and lung cancer: a case-control study of subjects participating in a mass-screening program. Chest. 2005;128(3):1490-6.
71. Sanikini H, Yuan JM, Butler LM, Koh WP, Gao YT, Steffen A, et al. Body mass index and lung cancer risk: a pooled analysis based on nested case-control studies from four cohort studies. BMC Cancer. 2018;18(1):220.
72. Tarleton HP, Park SL, Zhu WM, Lee YCA, Hashibe M, Morgenstern H, et al. Body mass index change in adulthood and lung and upper aerodigestive tract cancers. Int J Cancer. 2012;131(6):1407-16.
73. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med. 2008;27(8):1133-63.
74. Mariosa D, Carreras-Torres R, Martin RM, Johansson M, Brennan P. Commentary: What can Mendelian randomization tell us about causes of cancer? Int J Epidemiol. 2019;48(3):816-21.
75. Zhou W, Liu G, Hung RJ, Haycock PC, Aldrich MC, Andrew AS, et al. Causal relationships between body mass index, smoking and lung cancer: Univariable and multivariable Mendelian randomization. Int J Cancer. 2021;148(5):1077-86.
76. Wang T, Moon J-Y, Wu Y, Amos CI, Hung RJ, Tardon A, et al. Pleiotropy of genetic variants on obesity and smoking phenotypes: Results from the Oncoarray Project of The International Lung Cancer Consortium. PLoS One. 2017;12(9):e0185660.
77. Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. Int J Epidemiol. 2019;48(3):713-27.
78. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. Am J Epidemiol. 2015;181(4):251-60.

79. Carreras-Torres R, Johansson M, Haycock PC, Wade KH, Relton CL, Martin RM, et al. Obesity, metabolic factors and risk of different histological types of lung cancer: A Mendelian randomization study. PLoS One. 2017;12(6):e0177875.
80. Bey L, Hamilton MT. Suppression of skeletal muscle lipoprotein lipase activity during physical inactivity: a molecular reason to maintain daily low-intensity activity. J Physiol. 2003;551(2):673-82.
81. Hamilton MT, Hamilton DG, Zderic TW. Exercise physiology versus inactivity physiology: an essential concept for understanding lipoprotein lipase regulation. Exerc Sport Sci Rev. 2004;32(4):161.
82. Wilmot EG, Edwardson CL, Achana FA, Davies MJ, Gorely T, Gray LJ, et al. Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis. Springer; 2012.
83. Pedersen BK, Febbraio MA. Muscles, exercise and obesity: skeletal muscle as a secretory organ. Nat Rev Endocrinol. 2012;8(8):457.
84. Aoi W, Naito Y, Takagi T, Tanimura Y, Takanami Y, Kawai Y, et al. A novel myokine, secreted protein acidic and rich in cysteine (SPARC), suppresses colon tumorigenesis via regular exercise. Gut. 2012;62(6):882-9:gutjnl-2011-300776.
85. Hojman P, Dethlefsen C, Brandt C, Hansen J, Pedersen L, Pedersen BK. Exercise-induced muscle-derived cytokines inhibit mammary cancer cell growth. Am J Physiol Endocrinol Metab. 2011;301(3):E504-E10.
86. The Global Asthma Network. The Global Asthma Report 2018. 2018.
87. Azad N, Rojanasakul Y, Vallyathan V. Inflammation and lung cancer: roles of reactive oxygen/nitrogen species. J Toxicol Environ Health B Crit Rev. 2008;11(1):1-15.
88. Roche WR. Inflammatory and structural changes in the small airways in bronchial asthma. Am J Respir Crit Care Med. 1998;157(5 Pt 2):S191-4.
89. Zhang X, Liu Y, Shao H, Zheng X. Obesity Paradox in Lung Cancer Prognosis: Evolving Biological Insights and Clinical Implications. J Thorac Oncol. 2017;12(10):1478-88.
90. Khandekar MJ, Cohen P, Spiegelman BM. Molecular mechanisms of cancer development in obesity. Nat Rev Cancer. 2011;11(12):886-95.
91. Ribeiro R, Araújo A, Lopes C, Medeiros R. Immunoinflammatory mechanisms in lung cancer development: is leptin a mediator? J Thorac Oncol. 2007;2(2):105-8.
92. Petridou ET, Sergentanis TN, Antonopoulos CN, Dessypris N, Matsoukis IL, Aronis K, et al. Insulin resistance: an independent risk factor for lung cancer? Metabolism. 2011;60(8):1100-6.
93. Holmen J, Midthjell K, Krüger Ø, Langhammer A, Holmen TL, Bratberg GH, et al. The Nord-Trøndelag Health Study 1995–97 (HUNT 2): objectives, contents, methods and participation. Norsk epidemiologi. 2003;13(1):19-32.
94. Krokstad S, Langhammer A, Hveem K, Holmen T, Midthjell K, Stene T, et al. Cohort profile: the HUNT study, Norway. Int J Epidemiol. 2013;42(4):968-77.
95. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. GOLD Executive Summary. Am J Respir Crit Care Med. 2017;195(5):557-82.
96. Åsvold BO, Midthjell K, Krokstad S, Rangul V, Bauman A. Prolonged sitting may increase diabetes risk in physically inactive individuals: an 11 year follow-up of the HUNT Study, Norway. Diabetologia. 2017;60(5):830-5.
97. Chau JY, Grunseit A, Midthjell K, Holmen J, Holmen TL, Bauman AE, et al. Sedentary behaviour and risk of mortality from all-causes and cardiometabolic diseases in adults:

evidence from the HUNT3 population cohort. Br J Sports Med. 2013;49(11):737-42:bjsports-2012-091974.
98. Ekelund U, Steene-Johannessen J, Brown WJ, Fagerland MW, Owen N, Powell KE, et al. Does physical activity attenuate, or even eliminate, the detrimental association of sitting time with mortality? A harmonised meta-analysis of data from more than 1 million men and women. The Lancet. 2016;388(10051):1302-10.
99. Chau JY, Grunseit AC, Chey T, Stamatakis E, Brown WJ, Matthews CE, et al. Daily sitting time and all-cause mortality: a meta-analysis. PloS one. 2013;8(11):e80000.
100. Moe B, Eilertsen E, Nilsen TI. The Combined Effect of Leisure-Time Physical Activity and Diabetes on Cardiovascular Mortality The Nord-Trøndelag Health (HUNT) cohort study, Norway. Diabetes Care. 2013;36(3):690-5.
101. World Health Organization. Preventing and managing the global epidemic [report of a WHO Consultation on Obesity]. Geneva: World Health Organization; 2004.
102. Ferreira MA, Vonk JM, Baurecht H, Marenholz I, Tian C, Hoffman JD, et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. Nat Genet. 2017;49(12):1752-7.
103. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015;518(7538):197-206.
104. Cancer Registry of Norway. Cancer in Norway 2014-Cancer incidence, mortality, survival and prevalence in Norway. Oslo: Cancer Registry of Norway; 2015.
105. International classification of diseases for oncology (ICD-O). 3rd ed.: World Health Organization; 2013.
106. Larsen IK, Småstuen 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):1218-31.
107. Jiang L, Sun YQ, Brumpton BM, Langhammer A, Chen Y, Nilsen TIL, et al. Prolonged Sitting, Its Combination With Physical Inactivity and Incidence of Lung Cancer: Prospective Data From the HUNT Study. Front Oncol. 2019;9:101.
108. Sun YQ, Langhammer A, Wu C, Skorpen F, Chen Y, Nilsen TIL, et al. Associations of serum 25-hydroxyvitamin D level with incidence of lung cancer and histologic types in Norwegian adults: a case-cohort analysis of the HUNT study. Eur J Epidemiol. 2018;33(1):67-77.
109. Jiang L, Sun YQ, Langhammer A, Brumpton BM, Chen Y, Nilsen TI, et al. Asthma and asthma symptom control in relation to incidence of lung cancer in the HUNT study. Sci Rep. 2021;11(1):4539.
110. Staplin N, Herrington WG, Judge PK, Reith CA, Haynes R, Landray MJ, et al. Use of Causal Diagrams to Inform the Design and Interpretation of Observational Studies: An Example from the Study of Heart and Renal Protection (SHARP). Clin J Am Soc Nephrol. 2017;12(3):546-52.
111. Suttorp MM, Siegerink B, Jager KJ, Zoccali C, Dekker FW. Graphical presentation of confounding in directed acyclic graphs. Nephrol Dial Transplant. 2014;30(9):1418-23.
112. Kleinbaum DG, Klein M. Survival analysis: Springer; 2010.
113. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44(2):512-25.
114. Kantor ED, Hsu M, Du M, Signorello LB. Allergies and Asthma in Relation to Cancer Risk. Cancer Epidemiol Biomarkers Prev. 2019;28(8):1395-403.

115. Minelli C, Fabiola Del Greco M, van der Plaat DA, Bowden J, Sheehan NA, Thompson J. The use of two-sample methods for Mendelian randomization analyses on single large datasets. bioRxiv. 2020:2020.05.07.082206.
116. Barry C, Liu J, Richmond R, Rutter MK, Lawlor DA, Dudbridge F, et al. Exploiting collider bias to apply two-sample summary data Mendelian randomization methods to one-sample individual level data. medRxiv. 2020:2020.10.20.20216358.
117. Sanderson E, Windmeijer F. A weak instrument F-test in linear IV models with multiple endogenous variables. J Econom. 2016;190(2):212-21.
118. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet. 2018;50(5):693-8.
119. Cepelis A, Brumpton BM, Malmo V, Laugsand LE, Loennechen JP, Ellekjær H, et al. Associations of Asthma and Asthma Control With Atrial Fibrillation Risk: Results From the Nord-Trøndelag Health Study (HUNT). JAMA cardiology. 2018;3(8):721-8.
120. Laugsand LE, Strand LB, Platou C, Vatten LJ, Janszky I. Insomnia and the risk of incident heart failure: a population study. Eur Heart J. 2014;35(21):1382-93.
121. Heuch I, Heuch I, Hagen K, Mai X-M, Langhammer A, Zwart J-A. Is there an association between vitamin D status and risk of chronic low back pain? A nested case–control analysis in the Nord-Trøndelag Health Study. BMJ open. 2017;7(11):e018521.
122. Sun Y-Q, Langhammer A, Skorpen F, Chen Y, Mai X-M. Serum 25-hydroxyvitamin D level, chronic diseases and all-cause mortality in a population-based prospective cohort: the HUNT Study, Norway. BMJ open. 2017;7(6).
123. Lee KJ, Carlin JB. Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. Am J Epidemiol. 2010;171(5):624-32.
124. Lipsitch M, Tchetgen ET, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology (Cambridge, Mass). 2010;21(3):383.
125. López-Mesonero L, Márquez S, Parra P, Gamez-Leyva G, Munoz P, Pascual J. Smoking as a precipitating factor for migraine: a survey in medical students. J Headache Pain. 2009;10(2):101.
126. Taylor FR. Tobacco, Nicotine, and Headache. Headache. 2015;55(7):1028-44.
127. Le H, Tfelt-Hansen P, Skytthe A, Kyvik KO, Olesen J. Association between migraine, lifestyle and socioeconomic factors: a population-based cross-sectional study. J Headache Pain. 2011;12(2):157-72.
128. Textor J, van der Zander B, Gilthorpe MS, Liśkiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. Int J Epidemiol. 2016;45(6):1887-94.
129. Berry SD, Ngo L, Samelson EJ, Kiel DP. Competing Risk of Death: An Important Consideration in Studies of Older Adults. J Am Geriatr Soc. 2010;58(4):783-7.
130. Austin P, Fine J. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. Stat Med. 2017;36(27):4391-400.
131. Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. Int J Epidemiol. 2013;42(4):1134-44.
132. Sun Y-Q, Brumpton BM, Langhammer A, Chen Y, Kvaløy K, Mai X-M. Adiposity and asthma in adults: a bidirectional Mendelian randomisation analysis of the HUNT study. Thorax. 2020;75(3):202-8.
133. Parada H, McDonald E, Bellettiere J, Evenson KR, LaMonte MJ, LaCroix AZ. Associations of accelerometer-measured physical activity and physical activity-related cancer incidence in older women: results from the WHI OPACH Study. Br J Cancer. 2020;122(9):1409-16.

134. Baumeister SE, Leitzmann MF, Bahls M, Meisinger C, Amos CI, Hung RJ, et al. Physical Activity Does Not Lower the Risk of Lung Cancer. Cancer Res. 2020;80(17):3765-9.
135. Parimon T, Chien JW, Bryson CL, McDonell MB, Udris EM, Au DH. Inhaled corticosteroids and risk of lung cancer among patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2007;175(7):712-9.
136. Lee CH, Hyun MK, Jang EJ, Lee NR, Kim K, Yim JJ. Inhaled corticosteroid use and risks of lung cancer and laryngeal cancer. Respir Med. 2013;107(8):1222-33.
137. Gao C, Patel CJ, Michailidou K, Peters U, Gong J, Schildkraut J, et al. Mendelian randomization study of adiposity-related traits and risk of breast, ovarian, prostate, lung and colorectal cancer. Int J Epidemiol. 2016;45(3):896-908.
138. Rothman KJ, Greenland S, Lash TL. Modern epidemiology: Lippincott Williams & Wilkins; 2008.
139. Eekhout I, de Boer RM, Twisk JW, de Vet HC, Heymans MW. Missing data: a systematic review of how they are reported and handled. Epidemiology. 2012;23(5):729-32.
140. Kurtze N, Rangul V, Hustvedt B-E, Flanders WD. Reliability and validity of self-reported physical activity in the Nord-Trøndelag Health Study (HUNT 2). Eur J Epidemiol. 2007;22(6):379-87.
141. Toren K, Brisman J, Jarvholm B. Asthma and asthma-like symptoms in adults assessed by questionnaires. A literature review. Chest. 1993;104(2):600-8.
142. Langhammer A, Johnsen R, Holmen J, Gulsvik A, Bjermer L. Cigarette smoking gives more respiratory symptoms among women than among men The Nord-Trøndelag Health Study (HUNT). J Epidemiol Community Health. 2000;54(12):917-22.
143. Dahl R, Bjermer L. Nordic consensus report on asthma management. Respir Med. 2000;94(4):299-327.
144. Subramanian J, Govindan R. Lung cancer in never smokers: a review. J Clin Oncol. 2007;25(5):561-70.
145. Kommunekartlegginger-radon (Municipal Surveys-Radon). 2013 [19.April, 2018]. Available from: http://www.nrpa.no/temaartikler/90379/kart-og-data-radon.
146. Ulvestad B, Kjaerheim K, Møller B, Andersen A. Incidence trends of mesothelioma in Norway, 1965–1999. Int J Cancer. 2003;107(1):94-8.
147. Fewell Z, Davey Smith G, Sterne JAC. The Impact of Residual and Unmeasured Confounding in Epidemiologic Studies: A Simulation Study. Am J Epidemiol. 2007;166(6):646-55.
148. Burgess S, Swanson SA, Labrecque JA. Are Mendelian randomization investigations immune from bias due to reverse causation? Eur J Epidemiol. 2021;36(3):253-7.
149. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. Hum Mol Genet. 2018;27(R2):R195-R208.
150. Lippman SM, Klein EA, Goodman PJ, Lucia MS, Thompson IM, Ford LG, et al. Effect of selenium and vitamin E on risk of prostate cancer and other cancers: the Selenium and Vitamin E Cancer Prevention Trial (SELECT). JAMA. 2009;301(1):39-51.
151. Amderson DW, Mantel N. On epidemiologic surveys. Am J Epidemiol. 1983;118(5):613-9.

Paper I

frontiers in Oncology

ORIGINAL RESEARCH
published: 25 February 2019
doi: 10.3389/fonc.2019.00101

Prolonged Sitting, Its Combination With Physical Inactivity and Incidence of Lung Cancer: Prospective Data From the HUNT Study

Lin Jiang[1], Yi-Qian Sun[2], Ben Michael Brumpton[3,4,5], Arnulf Langhammer[1], Yue Chen[6], Tom I. L. Nilsen[1,7] and Xiao-Mei Mai[1]*

[1] Department of Public Health and Nursing, Faculty of Medicine and Health Science, Norwegian University of Science and Technology, Trondheim, Norway, [2] Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Science, Norwegian University of Science and Technology, Trondheim, Norway, [3] Department of Thoracic Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway, [4] K.G. Jebsen Centre for Genetic Epidemiology, Department of Public Health and Nursing, Norwegian University of Science and Technology, Trondheim, Norway, [5] MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom, [6] School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada, [7] Clinic of Anesthesia and Intensive Care, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway

OPEN ACCESS

Edited by:
Yawei Zhang,
Yale University, United States

Reviewed by:
Manuela Marron,
Leibniz Institute for Prevention Research and Epidemiology (LG), Germany
Xiangqian Guo,
Henan University, China

****Correspondence:***
Lin Jiang
lin.jiang@ntnu.no

Specialty section:
This article was submitted to Cancer Epidemiology and Prevention, a section of the journal Frontiers in Oncology

Received: *11 November 2018*
Accepted: *04 February 2019*
Published: *25 February 2019*

Citation:
Jiang L, Sun Y-Q, Brumpton BM, Langhammer A, Chen Y, Nilsen TIL and Mai X-M (2019) Prolonged Sitting, Its Combination With Physical Inactivity and Incidence of Lung Cancer: Prospective Data From the HUNT Study. Front. Oncol. 9:101. doi: 10.3389/fonc.2019.00101

Background: Prolonged sitting as a major sedentary behavior potentially contributes to illness, but its relation with lung cancer risk is unclear. Prolonged sitting can be presented in physically active or inactive individuals. Those who are extendedly seated and also physically inactive may represent the most sedentary people. We therefore aimed to prospectively examine if total sitting time daily itself or in combination with physical activity is associated with lung cancer incidence overall and histologic types.

Methods: We included 45,810 cancer-free adults who participated in the second survey of HUNT Study in Norway (1995–97), with a median follow-up of 18.3 years. Total sitting time daily and physical activity were self-reported at baseline. Lung cancer cases were ascertained from the Cancer Registry of Norway. Cox regression was used to estimate hazard ratios (HRs) with 95% confidence intervals (CIs).

Results: In total, 549 participants developed lung cancer during the follow-up. Total sitting time daily was not associated with the incidence of lung cancer overall and histologic subtypes. Compared with participants sitting <8 h daily and being physically active, those sitting ≥8 h daily (prolonged sitting) and being physically inactive had an increased incidence of lung cancer (overall: adjusted HR = 1.44, 95% CI: 1.07–1.94; small cell lung cancer: adjusted HR = 2.58, 95% CI: 1.23–5.41). Prolonged sitting only or physical inactivity only was not associated with the incidence of lung cancer.

Conclusions: Our study suggested that prolonged sitting was not independently associated with lung cancer incidence. The combination of prolonged sitting and physical inactivity might increase the risk of lung cancer. However, residual confounding by smoking cannot be excluded completely even though smoking was adjusted for with detailed information.

Keywords: prolonged sitting, physical inactivity, lung cancer risk, prospective cohort, HUNT study

INTRODUCTION

Lung cancer is one of the most common cancer types with a low survival rate (1). Small cell (SCLC) and non-small cell lung cancer (NSCLC) are the two major histologic types of lung cancer (2). Smoking is the most important risk factor for lung cancer, and less so for NSCLC than SCLC (3, 4). With a declining trend in smoking, other lifestyle factors may become more important for the incidence of lung cancer overall and histologic types. Physical activity or sedentary behavior has been suggested to be associated with the risk of cancer due to several plausible mechanisms including suppressed lipoprotein lipase activity (5, 6) and altered inflammatory pathways by lack of activities (7–9).

The relationship of physical activity with lung cancer risk has been extensively investigated. Recent meta-analysis studies have concluded that physical activity is associated with a reduced risk of lung cancer in smokers (10–12). Nonetheless, potential effects of physical activity and sedentary behavior might tangle in these meta-analysis studies since sedentary behavior was not properly taken into consideration in most of the individual studies.

Sedentary behavior describes a series of human behaviors requiring low energy expenditure in a sitting or reclining posture when awake (13). It is highly prevalent in western countries (14) and may be an independent risk factor for multiple health outcomes, including cancers (15, 16). Previous studies focused either on occupational sitting (17–19) or leisure-time TV watching (20, 21) in relation to lung cancer risk, with inconsistent results. Total sitting time daily is a better marker that reflects a sedentary lifestyle in the workplace, domestic environment and during leisure-time (22). However, there are limited studies on the relationship between total sitting time and lung cancer risk, and among them, one study found the association in a sub-population (23) and two studies did not adjust for physical activity properly due to lack of detailed information (21, 23). Physical inactivity and prolonged sitting are two distinct behaviors. Prolonged sitting can be present in physically active or inactive individuals. Those who are extendedly seated and also physically inactive may represent the most sedentary people. Thus, in the current study we aimed to prospectively examine the relationship between total sitting time daily and lung cancer risk (overall and major histologic types), taking smoking into consideration. We also investigated if different combinations of total sitting time and physical activity were associated with lung cancer incidence.

MATERIALS AND METHODS

Study Design and Population

The HUNT study is a large population-based health study in Norway, which includes more than 97% Caucasian participants and well-represents the Norwegian population. It consists of three consecutive surveys; HUNT1 (1984–1986), HUNT2 (1995-1997), and HUNT3 (2006–2008) (24). At each survey, all adults 20 years or older living in the area of Nord-Trøndelag were invited to complete extensive health and lifestyle questionnaires and undergo a clinical examination (25).

A total of 65,229 adults participated (70% of invited) in HUNT2. Every participant was followed-up from the date of participation in HUNT2 until the date of first diagnosis of lung cancer, the date of death or emigration from Norway or the end of follow-up on December 31, 2014, whichever came first. Diagnosis of lung cancer was obtained from the Cancer Registry of Norway. Information on vital status and emigration was obtained from the Central Population Registry.

Among the 65,229 participants, 59,070 self-reported no cancers at baseline. We included 45,810 cancer-free participants with complete information on total sitting time daily in the main cohort. Additionally, we investigated the combination of total sitting time and physical activity in relation to lung cancer risk in a sub-cohort of 33,793 participants who also provided complete information on physical activity.

The study was approved by the Regional Committee for Medical and Health Research Ethics of South-East Norway. All participants signed informed written consent on participation in HUNT, linkage to previous HUNT surveys and specific registries in accordance with the Declaration of Helsinki.

Ascertainment of Lung Cancer

By using the unique 11-digit personal identification number, participants' information from HUNT2 was linked to the Cancer Registry of Norway (26). Data from the Cancer Registry of Norway are reasonably accurate, complete (overall completeness 98.8% in 2001–05) and timely (27). The International Classification of Diseases version 10 (ICD-10) codes used for registration of lung cancer are C33-C34 (28). Histologic types were classified according to International Classification of Diseases of Oncology (ICD-O) (29).

Measurement of Exposures

Time spent sitting daily was measured by the question: "How many hours do you usually spend in the sitting position during a 24-h period? (work, meals, television, car etc.)." The participant reported the total number of hours as a positive integer. We categorized total sitting time daily into three categories (0–4, 5–7, and $\geq$8 h) based on similar cutoff criteria from former HUNT studies (30, 31) and two meta-analysis studies (32, 33). Occupational activity was used as an alternative marker for a sedentary lifestyle and measured by the question: "How would you describe your work?" Based on four response options, we categorized it into mostly sedentary work, much walking or lifting (two response options collapsed into one category), heavy physical work, and an "unknown" group with missing information.

Leisure-time physical activity was based on the question "How much of your leisure time have you been physically active per week during the last year?" Participants were asked to report average hours of light (no sweating or not being out of breath) and hard physical activity (sweating or out of breath) with the following response options for each intensity; none, $<$1 h, 1–2 h, and $\geq$3 h (reported as a positive integer). We classified participants' physical activity level as inactive (no any activity, or $\leq$2 h light activity only), low ($\geq$3 h light activity only, or $\leq$2 h light activity and $<$1 h hard activity), moderate ($\geq$3 h light

activity and <1 h hard activity, or 1–2 h hard activity regardless of light activity), and high (≥3 h hard activity regardless of light activity). This classification has demonstrated a dose-response relationship with mortality (34). Based on information of total sitting time and physical activity, we defined four combined categories: (1) total sitting time <8 h daily and physically active; (2) total sitting time <8 h daily and physically inactive; (3) total sitting time ≥8 h daily and physically active; and (4) total sitting time ≥8 h daily and physically inactive. Physically active referred to physical activity level from low to high. Physically inactive referred to no any activity or ≤2 h light activity only.

Information on Other Important Baseline Variables

Age, sex, body mass index (BMI), active smoking (status and pack-years), and passive smoking status, alcohol consumption, education, economic difficulties, family history of cancer and chronic bronchitis were included *a priori* as potential confounders. Weight and height in HUNT2 were measured by health professionals at clinical examination. BMI was calculated as weight in kilograms divided by height squared in meter (kg/m^2) and was grouped into three categories (<25.0, 25.0–29.9, and ≥30.0 kg/m^2) according to the recommendations of the World Health Organization (WHO) (35). Active smoking status was classified into never, former, current smokers and further classified based on pack-years (≤10.0, 10.1–20.0, and >20.1). Other variables were categorized as: passive smoking (never, only childhood, only adulthood, and both), alcohol consumption (never, 1–4, and ≥5 times/month), education (<10, 10–12, and ≥13 years, reported as a positive integer), economic difficulty (During the last year, has it at any time been difficulty to meet the cost of food, transportation, housing and such? yes/no), family history of cancer (Is there any family member such as father, mother, siblings who reported cancer? yes/no), and chronic bronchitis (Have you had a cough with phlegm for periods of at least 3 months during each of the last 2 years? yes/no). All missing information on the aforementioned variables was taken into analysis as an "unknown" category.

Statistical Analysis

Baseline characteristics of participants were presented by the categories of exposure variables. We used Cox proportional hazard models to examine the associations between total sitting time, and its combinations with physical activty and lung cancer incidence overall and histologic types and presented crude and adjusted hazard ratios (HR) with 95% confidence intervals (CI). Age was used as the time scale, with entry and exit time defined as the subject's age at recruitment and age at lung cancer diagnosis or censoring, respectively. When potential linear association between total sitting time and lung cancer risk was evaluated by likelihood ratio test, non-linearity was suggested ($p = 0.02$ for comparison between total sitting time used as a categorical variable and it used as an ordinal variable). Thus, the three categories (0–4, 5–7, and ≥8 h) of total sitting time were applied in the main analysis. Total sitting time was also categorized into tertiles.

In the sensitivity analysis, we first used occupational inactivity (mostly sedentary work) to test the robustness of our results on the assciation between total sitting time daily and lung cancer risk. Secondly, we performed analysis on the association of the combination groups of total sitting time and physcial activity with lung cancer risk after excluding the first three years of follow-up ($n = 33{,}322$) to reduce possible reverse causality by existing but undiagnosed lung cancer. In addition, we redefined the combination groups by including low activity into the physical inactivity level and repeated the analysis. Physical inactivity was thus redefined as no activity, any light activity only, or light activity ≤2 h and hard activity <1 h weekly. In this way, we could test if the original category of the combined sitting time ≥8 h and physical inactivity captured the most sedentary individuals. We further conducted complete case analysis among individuals with complete information on smoking (n=31,907) to minimize residual confounding from smoking. All statistical analyses were performed with STATA/SE 14.2 (College Station, TX, USA).

RESULTS

In total, 549 participants developed lung cancer during a median follow-up time of 18.3 years among the 45,810 subjects. **Tables 1**, **2** describe the distribution of baseline characteristics of participants according to total sitting time and its combination with physical activity levels. Compared to participants sitting <4 h or between 5 and 7 h daily, participants sitting ≥8 h were more likely to be males, frequent drinkers, have higher education and sedentary work (**Table 1**). Compared to participants who were sitting <8 h daily and physically active, people who were sitting ≥8 h daily and physically inactive were older, lower educated, more frequent smokers and sedentary workers. **Supplementary Table 1** shows that as compared to the original cancer-free population, participants in the main cohort had a higher percentage of family history of cancer and participants in the sub-cohort were relatively younger.

In **Table 3**, categories of total sitting time daily were not associated with lung cancer risk overall, SCLC or NSCLC in the main cohort after adjustment for a number of potential confounding factors including smoking status and physical activity. Total sitting time classified by tertiles was not associated with lung cancer risk either (data not presented). Results in ever smokers were similar to the main cohort (**Supplementary Table 2**). We were not able to perform analysis in never smokers as there were only 26 cases of lung cancer overall, no cases of SCLC and 19 cases of NSCLC among the never smokers.

Table 4 presents the association of the combined groups of total sitting time and physical activity with lung cancer risk overall and different histologic types. Compared to participants sitting <8 h and being physically active, participants sitting ≥8 h and being physically inactive had increased risks of lung cancer overall and SCLC in the main cohort (overall: adjusted HR = 1.44, 95% CI: 1.07–1.94; SCLC: adjusted HR = 2.58, 95% CI: 1.23–5.41). Neither of the group with prolonged

TABLE 1 | Baseline characteristics of participants according to total sitting time, the HUNT Study, 1995–97 ($N = 45,810$).

Variables	Total sitting time (hours*/day) 0–4 $N = 14,258$		5–7 $N = 14,549$		≥8 $N = 17,003$	
Age (years)	48.5	16.1	49.1	16.8	47.0	16.3
Body mass index (kg/m^2)	26.1	4.0	26.3	4.1	26.3	4.0
SEX						
Female	8,013	56.2	7,796	53.6	7,977	46.9
Male	6,245	43.8	6,753	46.4	9,026	53.1
SMOKING STATUS (PACK-YEARS)						
Never	6,181	43.4	6,075	41.8	7,477	44.0
Former ≤10.0	1,992	14.0	2,055	14.1	2,387	14.0
Former 10.1–20.0	730	5.1	858	5.9	981	5.8
Former >20.1	408	2.9	486	3.3	649	3.8
Current ≤10.0	1,568	11.0	1,551	10.7	1,708	10.1
Current 10.1–20.0	1,327	9.3	1,395	9.6	1,509	8.9
Current >20.1	855	6.0	1,022	7.0	1,237	7.3
Unknown	1,197	8.4	1,107	7.6	1,055	6.2
PASSIVE SMOKING STATUS						
Never	2,716	19.1	2,596	17.8	3,190	18.8
Only childhood	3,187	22.4	2,962	20.4	3,888	22.9
Only adulthood	2,364	16.6	2,446	16.8	2,467	14.5
Both childhood and adulthood	5,689	39.9	6,260	43.0	7,152	42.1
Unknown	302	2.1	285	2.0	306	1.8
PHYSICAL ACTIVITY						
Inactive	2,891	20.3	3,129	21.5	3,911	23.0
Low	2,467	17.3	2,672	18.4	3,331	19.6
Moderate	3,123	21.9	3,438	23.6	4,499	26.5
High	1,457	10.2	1,302	9.0	1,573	9.3
Unknown	4,320	30.3	4,008	27.6	3,689	21.7
ALCOHOL CONSUMPTION (TIMES/MONTH)						
Never	4,994	35.0	4,983	34.3	4,751	27.9
1-4	6,810	47.8	6,946	47.7	8,558	50.3
≥5	1,298	9.1	1,570	10.8	2,785	16.4
Unknown	1,156	8.1	1,050	7.2	909	5.4
EDUCATION (YEARS*)						
<10	5,145	36.1	5,170	35.5	4,063	23.9
10-12	5,281	37.0	5,133	35.3	5,281	31.1
≥13	3,376	23.7	3,783	26.0	7,246	42.6
Unknown	456	3.2	463	3.2	413	2.4
ECONOMIC DIFFICULTIES						
Yes	4,102	28.8	3,927	27.0	4,305	25.3
No	8,230	57.7	8,451	58.1	10,617	62.4
Unknown	1,926	13.5	2,171	14.9	2,081	12.2
FAMILY HISTORY OF CANCER						
Yes	4,171	29.3	4,382	30.1	4,888	28.8
No	10,087	70.8	10,167	69.9	12,115	70.7
CHRONIC BRONCHITIS						
Yes	433	3.0	471	3.2	550	3.2
No	13,561	95.1	13,811	94.9	16,182	95.2
Unknown	264	1.9	267	1.8	271	1.6
OCCUPATIONAL ACTIVITY						
Most sedentary work	839	5.9	1,736	11.9	8,706	51.2
Much walking or lifting at work	8,409	59.0	7,678	52.8	4,323	25.4
Heavy physical work	2,226	15.6	1,655	11.4	721	4.3
Unknown	2,784	19.5	3,480	23.9	3,253	19.1

*Continuous variables are presented with mean and standard deviation.Categorical variables are presented with number and column percentage of observations. *Hours of total sitting time daily and years of education were reported as a positive integer.*

sitting only or physical inactivity only was associated with lung cancer risk. Similar results were found among ever smokers (**Supplementary Table 3**).

In the sensitivity analysis, we found no association between occupational inactivity (mostly sedentary work) and lung cancer risk (**Supplementary Table 4**). The association of combined sitting ≥8 h and physical inactivity with lung cancer risk was not altered after excluding the first 3 years of follow-up (**Supplementary Table 5**). When grouping low level physical activity into the physical inactivity group (**Supplementary Table 6**), the associations of the combined sitting time ≥8 h and physical inactivity with the risks of lung cancer overall and histologic types became weaker compared to the original results. Complete case analysis for information of smoking showed comparable results for lung cancer overall (**Supplementary Table 7**).

DISCUSSION

Main Findings

In this prospective cohort study with 549 incident lung cancer cases, total sitting time daily was not associated with lung cancer. However, compared with participants sitting <8 h daily and being physically active, participants sitting ≥8 h daily and being physically inactive appeared to have increased risks of lung cancer overall and SCLC.

Comparison With Previous Studies

Previous studies showed different results on the association between sedentary lifestyle and lung cancer risk (17–21, 23). Occupational sitting was shown to be either protective (17, 19) or not associated with lung cancer risk (18). Different adjustment for physical activity and education might explain the differences in the results. On the contrary, leisure-time TV watching was associated with an increased risk of lung cancer among Japanese men but not women (20). Residual confounding by smoking was likely to be the explanation. In addition, Japanese women seemed to be more active than men (4.5 h housework for women & 1 h for men daily), which might

TABLE 2 | Baseline characteristics of participants according to combined total sitting time and physical activity level, the HUNT Study, 1995–97 ($N = 33,793$).

Total sitting time and physical activity level								
Variables	**Sitting <8 h &Physically active *N* = 14,459**		**Sitting <8 h & Physically inactive *N* = 6,020**		**Sitting ≥8 h & Physically active *N* = 9,403**		**Sitting ≥8 h& Physically inactive *N* = 3,911**	
Age (years)	44.2	15.3	49.4	16.5	42.2	14.0	49.8	17.2
Body mass index (kg/m^2)	25.8	3.8	26.7	4.4	25.8	3.7	26.8	4.5
SEX								
Female	7,215	49.9	3,616	60.1	4,030	42.9	2,018	51.6
Male	7,244	50.1	2,404	39.9	5,373	57.1	1,893	48.4
SMOKING STATUS (PACK-YEARS)								
Never	6,636	45.9	2,407	40.0	4,564	48.5	1,514	38.7
Former ≤10.0	2,283	15.8	829	13.8	1,469	15.6	517	13.2
Former 10.1–20.0	744	5.2	340	5.7	482	5.1	248	6.3
Former >20.1	371	2.8	213	3.5	275	2.9	179	4.6
Current ≤10.0	1,690	11.7	710	11.8	977	10.4	437	11.2
Current 10.1–20.0	1,158	8.0	671	11.2	713	7.6	402	10.3
Current >20.1	723	5.0	464	7.7	497	5.3	394	10.1
Unknown	854	5.9	386	6.4	426	4.5	220	5.6
PASSIVE SMOKING STATUS								
Never	2,844	19.7	1,016	16.9	1,890	20.1	667	17.1
Only childhood	3,600	24.9	1,208	20.1	2,532	26.9	747	19.1
Only adulthood	2,056	14.2	1,032	17.1	1,125	12.0	660	16.9
Both	5,775	39.9	2,692	44.7	3,745	39.8	1,787	45.7
Unknown	184	1.3	72	1.2	111	1.2	50	1.3
ALCOHOL CONSUMPTION (TIMES/MONTH)								
Never	3,980	27.5	2,398	39.8	1,904	20.3	1,385	35.4
1-4	7,855	54.3	2,744	45.6	5,282	56.2	1,814	46.4
≥5	1,808	12.5	496	8.2	1,864	19.8	521	13.3
Unknown	816	5.6	382	6.4	353	3.8	191	4.9
EDUCATION (YEARS*)								
<10	3,640	25.2	2,574	42.8	1,344	14.3	1,305	33.4
10-12	5,714	39.5	2,196	36.5	2,847	30.3	1,288	32.9
≥13	4,817	34.0	1,111	18.5	5,117	54.4	1,248	31.9
Unknown	188	1.3	139	2.3	95	1.0	70	1.8
ECONOMIC DIFFICULTIES								
Yes	4,295	29.7	1,843	30.6	2,474	26.3	1,057	27.0
No	8,964	62.0	3,266	54.3	6,453	68.6	2,187	55.9
Unknown	1,200	8.3	911	15.1	476	5.1	667	17.1
FAMILY HISTORY OF CANCER								
Yes	3,944	27.3	1,801	29.9	2,446	26.0	1,166	29.8
No	10,515	72.7	4,219	70.1	6,957	74.0	2,745	70.2
CHRONIC BRONCHITIS								
Yes	402	2.8	229	3.8	234	2.5	169	4.3
No	13,861	95.7	5,677	94.3	9,067	96.4	3,678	94.0
Unknown	196	1.4	114	1.9	102	1.1	64	1.6
OCCUPATIONAL ACTIVITY								
Most sedentary work	1,389	9.6	668	11.1	5,322	56.6	1,959	50.1
Much walking or lifting at work	8,966	62.0	3,385	56.2	2,650	28.2	938	24.0
Heavy physical work	2,225	15.4	771	12.8	397	4.2	159	4.0
Unknown	1,879	13.0	1,196	19.9	1,034	11.0	855	21.9

*Continuous variables are presented with mean and standard deviation. Categorical variables are presented with number and column percentage of observations. Physically active: physical activity level from low to high. Physically inactive: reported no activity or only light activity ≤2 h per week. *Years of education were reported as a positive integer.*

TABLE 3 | The association of total sitting time with lung cancer risk overall and histologic types, the HUNT Study, 1995–97 to 2014 ($N = 45,810$).

		Cases	IR (per 1000 person-years)	Crude HR	95% CI	Adjusted[a] HR	95% CI
LC overall	Sitting 0–4 h	185	0.76	1.00	Reference	1.00	Reference
	Sitting 5–7 h	165	0.67	0.86	0.70–1.06	0.82	0.66–1.01
	Sitting ≥8 h	199	0.69	1.09	0.89–1.33	1.05	0.66–1.29
SCLC	Sitting 0–4 h	25	0.10	1.00	Reference	1.00	Reference
	Sitting 5–7 h	20	0.08	0.78	0.43–1.40	0.73	0.40–1.31
	Sitting ≥8 h	31	0.11	1.28	0.75–2.17	1.18	0.69–2.03
NSCLC	Sitting 0–4 h	117	0.48	1.00	Reference	1.00	Reference
	Sitting 5–7 h	97	0.40	0.80	0.61–1.05	0.76	0.58–1.00
	Sitting ≥8 h	119	0.41	1.02	0.79–1.32	0.97	0.74–1.26

CI, Confidence interval; HR, Hazard ratio; IR, Incidence rate; LC, Lung cancer; NSCLC, Non-small cell lung cancer; SCLC, Small cell lung cancer. [a]Adjusted for sex, body mass index, smoking status (pack-years), passive smoking status, physical activity, alcohol consumption, education, economic difficulties, family history of cancer and chronic bronchitis. Age is used as the time scale.

TABLE 4 | The association of combined groups of total sitting time and physical activity with lung cancer risk overall and different histologic types, the HUNT Study, 1995–97 to 2014 ($N = 33,793$).

		Cases	IR (per 1,000 person-years)	Crude HR	95% CI	Adjusted[a]HR	95% CI
LC overall	Sitting <8 h & Physically active[b]	133	0.52	1.00	Reference	1.00	Reference
	Sitting <8 h & Physically inactive[c]	81	0.80	1.15	0.87–1.52	1.06	0.80–1.40
	Sitting ≥8 h & Physically active[b]	62	0.37	0.87	0.64–1.18	0.93	0.68–1.26
	Sitting ≥8 h & Physically inactive[c]	70	1.11	1.74	1.30–2.32	1.44	1.07–1.94
SCLC	Sitting <8 h & Physically active[b]	15	0.06	1.00	Reference	1.00	Reference
	Sitting <8 h & Physically inactive[c]	11	0.11	1.44	0.66–3.14	1.35	0.61–2.99
	Sitting ≥8 h & Physically active[b]	5	0.03	0.60	0.22–1.65	0.59	0.21–1.64
	Sitting ≥8 h & Physically inactive[c]	14	0.22	3.26	1.57–6.76	2.58	1.23–5.41
NSCLC	Sitting <8 h & Physically active[b]	78	0.31	1.00	Reference	1.00	Reference
	Sitting <8 h & Physically inactive[c]	54	0.54	1.34	0.95–1.90	1.21	0.85–1.72
	Sitting ≥8 h & Physically active[b]	41	0.25	0.96	0.65–1.40	1.01	0.68–1.48
	Sitting ≥8 h & Physically inactive[c]	39	0.62	1.68	1.15-2.48	1.36	0.92–2.01

CI, Confidence interval; HR, Hazard ratio; IR, Incidence rate; LC, Lung cancer; NSCLC, Non-small cell lung cancer; SCLC, Small cell lung cancer. [a]Adjusted for sex, body mass index, smoking status (pack-years), passive smoking status, alcohol consumption, education, economic difficulties, family history of cancer and chronic bronchitis. Age is used as the time scale. [b]Physically active: physical activity level from low to high. [c]Physically inactive: reported no activity or only light activity ≤2 h per week.

be another reason for the gender difference in lung cancer risk. However, no adjustment for physical activity was made in this study.

Total sitting time daily, as a better measure of sedentary lifestyle, was not associated with lung cancer risk in the current study. Our result was consistent with findings from the study by Wang et al. (23). We included both men and women and adjusted for levels of physical activity, whereas only menopausal women were included and no adjustment for physical activity was made in the referred study. In contrast, Lam et al. found a marginally increased risk of lung cancer associated with prolonged sitting among non-smokers (21). Although the cited study largely avoided confounding by smoking among non-smokers, the adjustment for physical activity only included vigorous activity. Leisure-time TV watching and occupational inactivity were also studied by Lam et al. but no associations with lung cancer risk were found. In our study, neither prolonged sitting nor occupational inactivity was independently associated with lung cancer incidence after adjustment for detailed categories of smoking and physical activity. However, we observed an increased risk of lung cancer among the most sedentary individuals who were both extendedly seated and physically inactive.

Possible Mechanisms

Although the underlying mechanisms on how the most sedentary individuals might have an increased risk of lung cancer are unclear, animal studies showed that lack of activities might suppress lipoprotein lipase activity in skeletal muscles and reduce glucose uptake (5, 6). Both are related to metabolic disorder that have been shown to be risk factors for several malignancies (6, 36). In addition, some pre-clinical studies suggest that weight-bearing skeletal muscles are not highly engaged during inactivity (7–9). This may alter anti-cancer responses of myokines in the skeletal muscles and activate inflammatory pathways that are important for cancer development (7–9).

Strengths and Limitations

Scientific evidence regarding total sitting time daily in relation to lung cancer risk overall and histologic types is scarce. Our study is the first prospective cohort study to investigate lung cancer risk among people who were both extendedly seated and physically inactive. In addition, the sample size of our study is relatively large with homogeneous study population and the follow-up period is long to allow for study of rare disease outcome such as lung cancer. The Cancer Registry of Norway records information about lung cancer diagnosis and different histologic types 1 year after the first diagnosis, and the information is soundly complete and accurate (26). The distribution of key baseline characteristics in both the main and sub-cohorts are generally similar to the original cancer-free population, suggesting no substantial selection bias. We also had information on a panel of potential confounders at baseline. In addition, we excluded participants with cancer at baseline in the main analysis and excluded the first 3 years of follow-up in the sensitivity analysis. Thus, potential reverse causation due to preexisting but undiagnosed lung cancer may not be a major problem.

However, our study has several limitations. First, misclassification of total sitting time and physical activity was possible due to self-reporting, and weak correlations with accelerometer counts ($r \approx 0.3$) were reported in a previous study (31). Since all information on exposures was collected at baseline before the diagnosis of lung cancer, it was more likely to be non-differential misclassification. We further used occupational inactivity as an alternative marker of sedentary lifestyle to test the robustness of the association between total sitting time and lung cancer risk, and similar results were observed. Additionally, in our sensitivity analysis, the magnitude of association of sitting time ≥ 8 h in combination with physical inactivity with the lung cancer risk was reduced by grouping individuals who had low level physical activity into the inactive group. This suggested that the original combination of prolonged sitting and physical inactivity identified the most sedentary individuals and thereby the highest risk group for lung cancer.

Second, individuals who were extendedly seated and physically inactive were more likely to be smokers than those who were shortly seated and physically active. To minimize confounding by smoking, we used detailed information on smoking status together with pack-years to categorize this variable, but we were not able to perform analysis in never smokers among which there were only 26 lung cancer cases. We further conducted complete case analysis for information on smoking and similar results were obtained. In addition, we redefined smoking status by including cessation years for former daily smokers and categorized subjects into the groups of never smokers, ex-smokers with smoking cessation >10.1 years, ex-smokers with smoking cessation ≤ 10.0 years, current smokers with 0–20.0 pack-years, and current smokers with >20.1 pack-years. The results were similar to our original findings (data not presented). Nevertheless, residual confounding by smoking cannot be excluded entirely. Other unmeasured factors such as hazardous occupational exposures might also confound the association. Indoor radon exposure is suggested to be the second important risk factor for lung cancer after smoking (37), but the level of indoor radon at any of the seven municipalities in Nord-Trøndelag was shown to be in the safety range (< 200 Bq/m3) (38). At last, we could not look into specific histologic types of NSCLC such as adenocarcinoma and squamous cell carcinoma due to limited statistical power.

In conclusion, our study suggested that prolonged sitting was not independently associated with lung cancer risk, but prolonged sitting in combination with physical inactivity might increase the risk of lung cancer. However, our results should be interpreted with caution due to a possibility of residual confounding of smoking.

ETHICS STATEMENT

The study was carried out in accordance with the recommendation of the Regional Committee for Medical and Health Research Ethics of South-East Norway. All subjects gave written informed consent on participation in HUNT, linkage to previous HUNT surveys and specific registries in accordance with the Declaration of Helsinki. The protocol was approved by the Regional Committee for Medical and Health Research Ethics of South-East Norway.

AUTHOR CONTRIBUTIONS

Y-QS and X-MM contributed to the study design and statistical analyses. AL and X-MM were responsible for data collection. LJ conducted statistical analyses, interpreted and wrote the initial draft of the manuscript. All authors participated in the data interpretation, contributed to the manuscript writing with important intellectual content and approved the final version of the manuscript.

FUNDING

LJ was supported by funding from the Extrastiftelsen through the Norwegian Cancer Society (Contract no. 77513001) and top-up funding from the collaboration partner between the Liaison Committee for Education, Research and Innovation in Central Norway and Central Norway Regional Health Authority. Y-QS was supported by the Norwegian Cancer Society (Project ID:5769155-2015).

ACKNOWLEDGMENTS

The HUNT Study is collaboration between HUNT Research Center (Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology—NTNU), Nord-Trøndelag County Council, Central Norway Regional Health Authority and the Norwegian Institute of Public Health.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.00101/full#supplementary-material

REFERENCES

1. International Agency for Research on Cancer. *GLOBOCAN 2012, Estimated cancer incidence, mortality and prevalence worldwide in 2012. 2016.* Available online at: http://globocan.iarc.fr/factsheet.asp
2. Travis WD, Harris C. *Pathology and Genetics of Tumours of the Lung, Pleura, Thymus and Heart*. IARC Press (2004).
3. Khuder SA. Effect of cigarette smoking on major histological types of lung cancer: a meta-analysis. *Lung Cancer* (2001) 31:139–48. doi: 10.1016/S0169-5002(00)00181-1
4. Lee PN, Forey BA, Coombs KJ. Systematic review with meta-analysis of the epidemiological evidence in the 1900s relating smoking to lung cancer. *BMC Cancer* (2012) 12:385. doi: 10.1186/1471-2407-12-385
5. Bey L, Hamilton MT. Suppression of skeletal muscle lipoprotein lipase activity during physical inactivity: a molecular reason to maintain daily low-intensity activity. *J Physiol.* (2003) 551:673–82. doi: 10.1113/jphysiol.2003.045591
6. Hamilton MT, Hamilton DG, Zderic TW. Exercise physiology versus inactivity physiology: an essential concept for understanding lipoprotein lipase regulation. *Exerc Sport Sci Rev*. (2004) 32:161. doi: 10.1097/00003677-200410000-00007
7. Pedersen BK, Febbraio MA. Muscles, exercise and obesity: skeletal muscle as a secretory organ. *Nat Rev Endocrinol.* (2012) 8:457. doi: 10.1038/nrendo.2012.49.
8. Hojman P, Dethlefsen C, Brandt C, Hansen J, Pedersen L, Pedersen BK. Exercise-induced muscle-derived cytokines inhibit mammary cancer cell growth. *Am J Physiol* (2011) 301:E504–10. doi: 10.1152/ajpendo.00520.2010
9. Aoi W, Naito Y, Takagi T, Tanimura Y, Takanami Y, Kawai Y, et al. A novel myokine, secreted protein acidic and rich in cysteine (SPARC), suppresses colon tumorigenesis via regular exercise. *Gut* (2012) 62:882–9. doi: 10.1136/gutjnl-2011-300776
10. Schmid D, Ricci C, Behrens G, Leitzmann MF. Does smoking influence the physical activity and lung cancer relation? A systematic review and meta-analysis. *Eur J Epidemiol.* (2016) 31:1173–90. doi: 10.1007/s10654-016-0186-y
11. Brenner DR, Yannitsos DH, Farris MS, Johansson M, Friedenreich CM. Leisure-time physical activity and lung cancer risk: a systematic review and meta-analysis. *Lung Cancer* (2016) 95:17–27. doi: 10.1016/j.lungcan.2016.01.021
12. Zhong S, Ma T, Chen L, Chen W, Lv M, Zhang X, et al. Physical activity and risk of lung cancer: a meta-analysis. *Clin J Sport Med.* (2016) 26:173–81. doi: 10.1097/JSM.0000000000000219
13. Barnes J, Behrens TK, Benden ME, Biddle S, Bond D, Brassard P, et al. Letter to the Editor: Standardized use of the terms" sedentary" and" sedentary behaviours". *Appl Physiol Nutr Metab.* (2012) 37:540–2. doi: 10.1139/h2012-024
14. Matthews CE, Chen KY, Freedson PS, Buchowski MS, Beech BM, Pate RR, et al. Amount of time spent in sedentary behaviors in the United States, 2003–2004. *Am J Epidemiol.* (2008) 167:875–81. doi: 10.1093/aje/kwm390
15. Thorp AA, Owen N, Neuhaus M, Dunstan DW. Sedentary behaviors and subsequent health outcomes in adults: a systematic review of longitudinal studies, 1996–2011. *Am J Prevent Med.* (2011) 41:207–15. doi: 10.1016/j.amepre.2011.05.004
16. Shen D, Mao W, Liu T, Lin Q, Lu X, Wang Q, et al. Sedentary behavior and incident cancer: a meta-analysis of prospective studies. *PLoS ONE* (2014) 9:e105709. doi: 10.1371/journal.pone.0105709
17. 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) 119:2389–97. doi: 10.1002/ijc.22125
18. Thune I, Lund E. The influence of physical activity on lung-cancer risk: a prospective study of 81,516 men and women. *Int J Cancer* (1997) 70:57–62.
19. Bak H, Christensen J, Thomsen BL, Tjønneland A, Overvad K, Loft S, et al. Physical activity and risk for lung cancer in a Danish cohort. *Int J Cancer* (2005) 116:439–44. doi: 10.1002/ijc.21085
20. Ukawa S, Tamakoshi A, Wakai K, Noda H, Ando M, Iso H. Prospective cohort study on television viewing time and incidence of lung cancer: findings from the Japan Collaborative Cohort Study. *Cancer Causes Control* (2013) 24:1547–53. doi: 10.1007/s10552-013-0231-z
21. Lam TK, Moore SC, Brinton LA, Smith L, Hollenbeck AR, Gierach GL, et al. Anthropometric measures and physical activity and the risk of lung cancer in never-smokers: a prospective cohort study. *PLoS ONE* (2013) 8:e70672. doi: 10.1371/journal.pone.0070672
22. Katzmarzyk PT, Church TS, Craig CL, Bouchard C. Sitting time and mortality from all causes, cardiovascular disease, and cancer. *Med Sci Sports Exerc.* (2009) 41:998–1005. doi: 10.1249/MSS.0b013e3181930355
23. Wang A, Qin F, Hedlin H, Desai M, Chlebowski R, Gomez S, et al. Physical activity and sedentary behavior in relation to lung cancer incidence and mortality in older women: The Women's Health Initiative. *Int J Cancer*. (2016) 139:2178–92. doi: 10.1002/ijc.30281
24. Holmen J, Midthjell K, Krüger Ø, Langhammer A, Holmen TL, Bratberg GH, et al. The Nord-Trøndelag Health Study 1995–97 (HUNT 2): objectives, contents, methods and participation. *Norsk Epidemiologi*. (2003) 13:19–32. doi: 10.5324/nje.v13i1.305
25. Krokstad S, Langhammer A, Hveem K, Holmen T, Midthjell K, Stene T, et al. Cohort profile: the HUNT study, Norway. *Int J Epidemiol*. (2013) 42:968–77. doi: 10.1093/ije/dys095
26. Cancer Registry of Norway. (2017). Available online: http://www.kreftregisteret.no/
27. Larsen IK, Småstuen 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) 45:1218–31. doi: 10.1016/j.ejca.2008.10.037
28. Cancer Registry of Norway. *Cancer in Norway 2014-Cancer Incidence, Mortality, Survival and Prevalence in Norway*. Oslo: Cancer Registry of Norway (2015).
29. ICD-O. *International Classification of Diseases for Oncology (ICD-O)*. 3rd ed. World Health Organization (2013).
30. Åsvold BO, Midthjell K, Krokstad S, Rangul V, Bauman A. Prolonged sitting may increase diabetes risk in physically inactive individuals: an 11 year follow-up of the HUNT Study, Norway. *Diabetologia* (2017) 60:830–5. doi: 10.1007/s00125-016-4193-z
31. Chau JY, Grunseit A, Midthjell K, Holmen J, Holmen TL, Bauman AE, et al. Sedentary behaviour and risk of mortality from all-causes and cardiometabolic diseases in adults: evidence from the HUNT3 population cohort. *Br J Sports Med*. (2013) 49:737–42. doi: 10.1136/bjsports-2012-091974
32. Ekelund U, Steene-Johannessen J, Brown WJ, Fagerland MW, Owen N, Powell KE, et al. Does physical activity attenuate, or even eliminate, the detrimental association of sitting time with mortality? A harmonised meta-analysis of data from more than 1 million men and women. *Lancet* (2016) 388:1302–10. doi: 10.1016/S0140-6736(16)30370-1
33. Chau JY, Grunseit AC, Chey T, Stamatakis E, Brown WJ, Matthews CE, et al. Daily sitting time and all-cause mortality: a meta-analysis. *PLoS ONE* (2013) 8:e80000. doi: 10.1371/journal.pone.0080000
34. Moe B, Eilertsen E, Nilsen TI. The Combined Effect of Leisure-Time Physical Activity and Diabetes on Cardiovascular Mortality The Nord-Trøndelag Health (HUNT) cohort study, Norway. *Diabetes Care* (2013) 36:690–5. doi: 10.2337/dc11-2472
35. WHO. *Preventing and Managing the Global Epidemic*. Report of a WHO Consultation on Obesity. Geneva: World Health Organization (2004).
36. Wilmot EG, Edwardson CL, Achana FA, Davies MJ, Gorely T, Gray LJ, et al. Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis. *Diabetologia* (2012) 55:2895–905. doi: 10.1007/s00125-012-2677-z
37. WHO. *WHO Handbook on Indoor Radon: A Public Health Perspective*. Geneva: World Health Organization (2009).
38. *Kommunekartlegginger-radon (Municipal Surveys-Radon)*. (2013) Available online at: http://www.nrpa.no/temaartikler/90379/kart-og-data-radon (Accessed April 19, 2018).

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary material: Paper I

Supplementary table 1. Comparison of baseline characteristics of participants in the main and sub-cohorts with the original cancer-free population, the HUNT Study, 1995-97

Variables	Original cancer-free population (N = 59,070)		Main Cohort (N = 45,810)		Sub-Cohort (N = 33,793)	
Age (years)	48.7	16.8	48.1	16.4	45.2	15.7
Body mass index (kg/m^2)	26.3	4.1	26.2	4.0	26.1	4.0
Sex						
Female	30,819	52.2	23,786	51.9	16,879	50.0
Male	28,251	47.8	22,024	48.1	16,914	50.0
Smoking status (pack-years)						
Never	25,299	42.8	19,733	43.1	15,121	44.8
Former ≤10.0	7826	13.3	6434	14.0	5098	15.1
Former 10.1-20.0	3116	5.3	2569	5.6	1814	5.4
Former >20.1	1911	3.2	1543	3.4	1038	3.1
Current ≤10.1	6366	10.8	4827	10.5	3814	11.3
Current 10.1-20.0	5638	9.5	4231	9.2	2944	8.7
Current >20.1	4087	6.9	3114	6.8	2078	6.2
Unknown	4827	8.2	3359	7.3	1886	5.6
Passive smoking status						
Never	10,889	18.4	8502	18.6	6417	19.0
Only childhood	12,587	21.3	10,037	21.9	8087	23.9
Only adulthood	9539	16.2	7277	15.9	4873	14.4
Both	24,640	41.7	19,101	41.7	13,999	41.4
Unknown	1415	2.4	893	2.0	417	1.2
Alcohol consumption (times/month)						
Never	19,910	33.7	14,728	32.2	9667	28.6
1-4	27,663	46.8	22,314	48.7	17,695	52.4
≥5	6907	11.7	5653	12.3	4689	13.9
Unknown	4590	7.8	3115	6.8	1742	5.2
Education (years)						
<10	19,652	33.3	14,378	31.4	8863	26.2
10-12	19,672	33.3	15,695	34.3	12,045	35.6
≥13	17,428	29.5	14,405	31.5	12,393	36.7
Unknown	2318	3.9	1332	2.9	492	1.5
Economic difficulties						
Yes	12,974	22.0	12,334	26.9	9669	28.6
No	29,065	49.2	27,298	59.6	20,870	61.8
Unknown	17,031	28.8	6178	13.5	3254	9.6
Family history of cancer						
Yes	14,757	25.0	13,441	29.3	9357	27.7
No	44,313	75.0	32,369	70.7	24,436	72.3
Chronic bronchitis						
Yes	1914	3.2	1454	3.2	1034	3.1
No	56,054	94.9	43,554	95.1	32,283	95.5
Unknown	1102	1.9	802	1.8	476	1.4
Occupational activity						
Most sedentary work	13,835	23.5	11,281	24.6	9338	27.6
Much walking or lifting at work	25,722	43.5	20,410	44.5	15,939	47.2
Heavy physical work	5911	10.0	4602	10.1	3552	10.5
Unknown	13,602	23.0	9517	20.8	4964	14.7

Abbreviations: HUNT, Nord-Trøndelag Health Study; SD, standard deviation. Continuous variables are presented with mean and standard deviation.Categorical variables are presented with number and column percentage of observations.

Supplementary table 2. The association of total sitting time with lung cancer risk overall and histologic types in ever smokers, the HUNT Study, 1995-97 to 2014 (N = 25,449)

		N	Cases	IR (per 1000 person-years)	Crude HR	Adjusted[1] HR	95% CI
LC overall	Sitting 0-4 hours	7832	171	1.30	1.00	1.00	Reference
	Sitting 5-7 hours	8284	155	1.13	0.85	0.83	0.67–1.03
	Sitting ≥8 hours	9333	189	1.21	1.09	1.07	0.87–1.33
SCLC	Sitting 0-4 hours	7832	24	0.18	1.00	1.00	Reference
	Sitting 5-7 hours	8284	20	0.15	0.78	0.75	0.42–1.37
	Sitting ≥8 hours	9333	31	0.20	1.27	1.23	0.71–2.14
NSCLC	Sitting 0-4 hours	7832	108	0.82	1.00	1.00	Reference
	Sitting 5-7 hours	8284	90	0.65	0.78	0.76	0.58–1.01
	Sitting ≥8 hours	9333	111	0.71	1.00	0.97	0.74–1.27

Abbreviation: CI, Confidence interval; HR, Hazard ratio; HUNT, Nord-Trøndelag Health Study; IR, Incidence rate; LC, Lung cancer; NSCLC, Non-small cell lung cancer; SCLC, Small cell lung cancer. [1] Adjusted for sex, body mass index, smoking, passive smoking, physical activity, alcohol consumption, education, economic difficulties, family history of cancer and chronic bronchitis. Age is used as the time scale.

Supplementary table 3. The association of combined groups of total sitting time and physical activity with lung cancer risk overall and different histologic types in ever smokers, the HUNT Study, 1995-97 to 2014 (N = 18,335)

		N	Cases	IR (per 1000 person-years)	Crude HR	Adjusted[1] HR	95% CI
LC overall	Sitting <8 hours &physically active[2]	7668	118	0.89	1.00	1.00	Reference
	Sitting <8 hours &physically inactive[3]	3541	76	1.30	1.26	1.13	0.85–1.52
	Sitting ≥8 hours &physically active[2]	4764	58	0.70	0.91	0.99	0.71–1.36
	Sitting ≥8 hours &physically inactive[3]	2362	69	1.80	1.93	1.61	1.19–2.18
SCLC	Sitting <8 hours &physically active[2]	7668	15	0.11	1.00	1.00	Reference
	Sitting <8 hours &physically inactive[3]	3541	10	0.17	1.30	1.22	0.54–2.77
	Sitting ≥8 hours &physically active[2]	4764	5	0.06	0.60	0.58	0.21–1.63
	Sitting ≥8 hours &physically inactive[3]	2362	14	0.37	3.06	2.55	1.22–5.36
NSCLC	Sitting <8 hours &physically active[2]	7668	67	0.51	1.00	1.00	Reference
	Sitting <8 hours &physically inactive[3]	3541	51	0.87	1.49	1.34	0.92–1.94
	Sitting ≥8 hours &physically active[2]	4764	38	0.46	1.03	1.09	0.73–1.64
	Sitting ≥8 hours &physically inactive[3]	2362	39	1.02	1.91	1.58	1.05–2.36

Abbreviation: CI, Confidence interval; HR, Hazard ratio; HUNT, Nord-Trøndelag Health Study; IR, Incidence rate; LC, Lung cancer; NSCLC, Non-small cell lung cancer; SCLC, Small cell lung cancer. [1] Adjusted for sex, body mass index, smoking, passive smoking, alcohol consumption, education, economic difficulties, family history of cancer and chronic bronchitis Age is used as the time scale.[2] Physically active: physical activity level from low to high. [3] Physically inactive: reported no activity or only light activity ≤2 hours per week.

Supplementary table 4. The association of occupational inactivity with lung cancer risk overall and histologic types, the HUNT Study, 1995-97 to 2014 (N = 45,810)

		N	Cases	IR (per 1000 person-years)	Crude HR	Adjusted[1] HR	95% CI
LC overall	Much walking or lifting at work	20,410	189	0.52	1.00	1.00	Reference
	Most sedentary work	11,281	102	0.52	0.92	0.97	0.75–1.24
	Heavy physical work	4602	47	0.58	0.98	0.79	0.57–1.10
	Unknown	9517	211	1.55	1.13	1.09	0.87–1.36
SCLC	Much walking or lifting at work	20,410	28	0.08	1.00	1.00	Reference
	Most sedentary work	11,281	13	0.07	0.81	0.79	0.40–1.55
	Heavy physical work	4602	2	0.02	0.28	0.23	0.05–0.98
	Unknown	9517	33	0.24	1.45	1.35	0.77–2.37
NSCLC	Much walking or lifting at work	20,410	121	0.33	1.00	1.00	Reference
	Most sedentary work	11,281	61	0.31	0.87	0.90	0.66–1.23
	Heavy physical work	4602	33	0.41	1.08	0.87	0.58–1.29
	Unknown	9517	118	0.87	1.07	1.02	0.77–1.36

Abbreviation: CI, Confidence interval; HR, Hazard ratio; HUNT, Nord-Trøndelag Health Study; IR, Incidence rate; LC, Lung cancer; NSCLC, Non-small cell lung cancer; SCLC, Small cell lung cancer. [1] Adjusted for sex, body mass index, smoking, passive smoking, physical activity, alcohol consumption, education, economic difficulties, family history of cancer and chronic bronchitis. Age is used as the time scale.

Supplementary table 5. The association of combined groups of total sitting time and physical activity with lung cancer risk overall and different histologic types after excluding the first three years' follow-up for all participants, the HUNT Study, 1995-97 to 2014 (N = 33,322)

		N	Cases	IR (per 1000 person-years)	Crude HR	Adjusted[1] HR	95% CI
LC overall	Sitting <8 hours & physically active[2]	14,321	117	0.56	1.00	1.00	Reference
	Sitting <8 hours & physically inactive[3]	5898	74	0.89	1.22	1.09	0.82–1.47
	Sitting ≥8 hours & physically active[2]	9322	52	0.38	0.82	0.87	0.63–1.22
	Sitting ≥8 hours & physically inactive[3]	3781	62	1.20	1.82	1.47	1.07–2.00
SCLC	Sitting <8 hours & physically active[2]	14,321	15	0.07	1.00	1.00	Reference
	Sitting <8 hours & physically inactive[3]	5898	10	0.12	1.34	1.22	0.54–2.76
	Sitting ≥8 hours & physically active[2]	9322	4	0.03	0.47	0.45	0.15–1.38
	Sitting ≥8 hours & physically inactive[3]	3781	13	0.25	3.13	2.38	1.12–5.08
NSCLC	Sitting <8 hours & physically active[2]	14,321	68	0.32	1.00	1.00	Reference
	Sitting <8 hours & physically inactive[3]	5898	48	0.58	1.39	1.21	0.83–1.77
	Sitting ≥8 hours & physically active[2]	9322	36	0.26	0.95	0.99	0.66–1.50
	Sitting ≥8 hours & physically inactive[3]	3781	34	0.66	1.74	1.36	0.90–2.07

Abbreviation: CI, Confidence interval; HR, Hazard ratio; HUNT, Nord-Trøndelag Health Study; IR, Incidence rate; LC, Lung cancer; NSCLC, Non-small cell lung cancer; SCLC, Small cell lung cancer. [1] Adjusted for sex, body mass index, smoking, passive smoking, alcohol consumption, education, economic difficulties, family history of cancer and chronic bronchitis. Age is used as the time scale.[2] physically active: physical activity level from low to high. [3] Physically inactive: reported no activity or only light activity ≤2 hours per week.

Supplementary table 6. The association of combined groups of total sitting time and physical activity with lung cancer risk overall and different histologic types after including low activity into the physically inactive group, the HUNT Study, 1995-97 to 2014 (N = 33,793)

		N	Cases	IR (per 1000 person-years)	Crude HR	Adjusted[1] HR	95% CI
LC overall	Sitting <8 hours & physically active[2]	9320	70	0.42	1.00	1.00	Reference
	Sitting <8 hours & physically inactive[3]	11,159	144	0.76	1.27	1.16	0.87–1.55
	Sitting ≥8 hours & physically active[2]	6072	30	0.28	0.79	0.89	0.58–1.37
	Sitting ≥8 hours & physically inactive[3]	7242	102	0.84	1.63	1.40	1.03–1.90
SCLC	Sitting <8 hours & physically active[2]	9320	8	0.05	1.00	1.00	Reference
	Sitting <8 hours & physically inactive[3]	11,159	18	0.10	1.44	1.34	0.57–3.12
	Sitting ≥8 hours & physically active[2]	6072	2	0.02	0.44	0.47	0.10–2.24
	Sitting ≥8 hours & physically inactive[3]	7242	17	0.14	2.47	1.99	0.85–4.67
NSCLC	Sitting <8 hours & physically active[2]	9320	43	0.26	1.00	1.00	Reference
	Sitting <8 hours & physically inactive[3]	11,159	89	0.47	1.31	1.17	0.81–1.70
	Sitting ≥8 hours & physically active[2]	6072	22	0.20	0.92	1.01	0.60–1.71
	Sitting ≥8 hours & physically inactive[3]	7242	58	0.48	1.53	1.27	0.85–1.90

Abbreviation: CI, Confidence interval; HR, Hazard ratio; HUNT, Nord-Trøndelag Health Study; IR, Incidence rate; LC, Lung cancer; NSCLC, Non-small cell lung cancer; SCLC, Small cell lung cancer. [1] Adjusted for sex, body mass index, smoking, passive smoking, alcohol consumption, education, economic difficulties, family history of cancer and chronic bronchitis. Age is used as the time scale. [2] Physically active: physical activity level from moderate to high. [3] Physically inactive: reported no activity or any light activity only or light activity ≤2 hours and hard activity <1 hour per week.

Supplementary table 7. Complete case analysis regarding information on smoking status for the association of combined groups of total sitting time and physical activity with lung cancer risk overall and different histologic types, the HUNT Study, 1995-97 to 2014 (N = 31,907)

		N	Cases	IR (per 1000 person-years)	Crude HR	Adjusted[1] HR	95% CI
LC overall	Sitting <8 hours & physically active[2]	13,605	116	0.48	1.00	1.00	Reference
	Sitting <8 hours & physically inactive[3]	5634	73	0.77	1.19	1.05	0.78–1.41
	Sitting ≥8 hours & physically active[2]	8977	58	0.36	0.90	0.97	0.70–1.33
	Sitting ≥8 hours & physically inactive[3]	3691	64	1.06	1.79	1.44	1.06–1.97
SCLC	Sitting <8 hours & physically active[2]	13,605	15	0.06	1.00	1.00	Reference
	Sitting <8 hours & physically inactive[3]	5634	10	0.11	1.30	1.21	0.53–2.75
	Sitting ≥8 hours & physically active[2]	8977	5	0.03	0.59	0.57	0.20–1.59
	Sitting ≥8 hours & physically inactive[3]	3691	10	0.17	2.27	1.72	0.77–3.88
NSCLC	Sitting <8 hours & physically active[2]	13,605	70	0.29	1.00	1.00	Reference
	Sitting <8 hours & physically inactive[3]	5634	47	0.50	1.30	1.14	0.78–1.66
	Sitting ≥8 hours & physically active[2]	8977	39	0.24	0.98	1.03	0.69–1.54
	Sitting ≥8 hours & physically inactive[3]	3691	37	0.62	1.75	1.40	0.94–2.10

Abbreviation: CI, Confidence interval; HR, Hazard ratio; HUNT, Nord-Trøndelag Health Study; IR, Incidence rate; LC, Lung cancer; NSCLC, Non-small cell lung cancer; SCLC, Small cell lung cancer. [1] Adjusted for sex, body mass index, smoking, passive smoking, alcohol consumption, education, economic difficulties, family history of cancer and chronic bronchitis. Age is used as the time scale.[2] Physically active: physical activity level from low to high. [3] Physically inactive: reported no activity or only light activity ≤2 hours per week.

Paper II

scientific reports

Check for updates

OPEN

Asthma and asthma symptom control in relation to incidence of lung cancer in the HUNT study

Lin Jiang[1]✉, Yi-Qian Sun[2,3,4], Arnulf Langhammer[5,6], Ben Michael Brumpton[7,8,9], Yue Chen[10], Tom IL Nilsen[1,11], Linda Leivseth[12], Sissel Gyrid Freim Wahl[2,3] & Xiao-Mei Mai[1]

Large prospective studies on asthma, especially asthma symptom control, as a potential risk factor for lung cancer are limited. We followed up 62,791 cancer-free Norwegian adults from 1995–1997 to 2017. Self-reported doctor-diagnosed asthma was categorized into active and non-active asthma. Levels of asthma symptom control were classified into controlled and partially controlled (including partly controlled and uncontrolled) according to the Global Initiative for Asthma guidelines. Incident lung cancer cases were ascertained from the Cancer Registry of Norway. Cox regression models were used to estimate hazard ratios (HRs) with 95% confidence intervals (CIs) for possible associations. Totally, 984 participants developed lung cancer during a median follow-up of 21.1 years. After adjustment for smoking and other potential confounders, an increased incidence of lung cancer was found for adults with partially controlled asthma (HR 1.39, 95% CI 1.00–1.92) compared with those without asthma at baseline. Adults with active asthma had a tendency of increased lung cancer incidence (HR 1.29, 95% CI 0.95–1.75). Sensitivity analyses indicated that the observed associations were less likely resulted from reverse causation or residual confounding by smoking. Our findings suggested that proper control of asthma symptoms might contribute to a reduced incidence of lung cancer.

Asthma is a common chronic lung disease characterized by chronic inflammation, reversible airway obstruction and enhanced bronchial reactivity[1,2]. The chronic inflammatory state of the lung among those with asthma has been suggested to cause oxidative injuries that may lead to lung cancer development[3].

Given the relatively high prevalence of asthma and low survival rate of lung cancer worldwide, it is important to clarify if asthma is a risk factor for lung cancer. So far, three meta-analysis studies have investigated the possible association between asthma and lung cancer and a positive association has been suggested[2,4,5]. More case–control studies than prospective studies were included in the meta-analysis studies. Among the included prospective studies, several used age and sex-standardized incidence ratio (SIR) to present the relative risk for lung cancer in patients with asthma compared with a general population[6–8]. Smoking, the most important confounding factor, was not controlled for. Also, the included prospective cohort studies tended to have inadequate number of lung cancer cases due to short follow-up duration (< 10 years)[9,10] or small sample size[11,12]. A recent prospective cohort study using data from the UK Million Women Study with more than 14 years of follow-up reported that asthma requiring treatment was associated with an increased incidence of lung cancer in never smoking women[13]; however, it investigated 34 potential risk factors simultaneously and a chance finding could not be excluded.

[1]Department of Public Health and Nursing, Faculty of Medicine and Health Science, Norwegian University of Science and Technology (NTNU), Postbox 8905, MTFS, N-7491, Trondheim, Norway. [2]Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Science, Norwegian University of Science and Technology, Trondheim, Norway. [3]Department of Pathology, Clinic of Laboratory Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway. [4]TkMidt-Center for Oral Health Services and Research, Mid-Norway, Trondheim, Norway. [5]HUNT Research Centre, Department of Public Health and Nursing, Norwegian University of Science and Technology, Levanger, Norway. [6]Levanger Hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway. [7]Clinic of Thoracic and Occupational Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway. [8]K.G. Jebsen Centre for Genetic Epidemiology, Department of Public Health and Nursing, Norwegian University of Science and Technology, Trondheim, Norway. [9]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. [10]School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, Canada. [11]Clinic of Anesthesia and Intensive Care, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway. [12]Centre for Clinical Documentation and Evaluation (SKDE), Northern Norway Regional Health Authority, Tromsø, Norway. ✉email: lin.jiang@ntnu.no

Variables	No asthma	Non-active asthma	Active asthma
Number of subjects	**59,591**	**1090**	**2110**
Age (years)	49.5 ± 17.0	47.3 ± 17.2	51.1 ± 17.3
Body mass index (kg/m^2)	26.3 ± 4.0	26.7 ± 4.4	27.4 ± 4.9
Number of lung cancer cases (%)	**921 (1.6)**	**14 (1.3)**	**49 (2.3)**
Sex, % (women/men)	53.0/47.0	51.2/48.8	56.2/43.8
Allergic rhinitis, % (no/yes/unknown)	70.5/5.5/24.0	40.2/45.8/14.0	35.6/50.0/14.4
Smoking status, % (never/current/former/unknown)	43.0/28.6/26.3/2.2	40.6/28.2/29.0/2.2	37.3/28.5/32.0/2.1
Passive smoking, % (never/ever/unknown)	18.5/79.6/1.9	15.9/83.0/1.1	14.7/83.2/2.0
Alcohol consumption (times/month), % (never/ ≥ 1/unknown)	34.6/56.9/8.6	33.9/58.0/8.2	39.0/52.4/8.2
Physical activity, % (inactive[1]/active[2]/unknown)	21.6/48.0/30.4	25.4/50.4/24.2	22.4/45.9/31.7
Total sitting time daily (hours), % (< 8/ ≥ 8/unknown)	48.1/27.8/24.1	52.2/35.2/12.6	54.6/31.9/13.6
Education (years), % (< 10/ ≥ 10/unknown)	34.0/61.0/5.1	31.8/63.1/5.1	38.5/55.5/6.0
Economic difficulties, % (no/yes/unknown)	48.0/21.1/30.9	51.5/28.1/20.5	47.4/31.1/21.5
Family history of cancer, % (no/yes)	75.0/25.0	72.8/27.2	69.1/30.9

Table 1. Distribution of baseline characteristics according to asthma categories in the HUNT2 Study, 1995–1997 (n = 62,791). HUNT: Nord-Trøndelag Health Study. Data are given as mean ± standard deviation or percentage of subjects in each asthma category. [1]Inactive: no physical activity or only light physical activity ≤ 2 h per week. [2]Active: physical activity level from low to high.

Asthma is not curable, but it is controllable with appropriate treatment. Nonetheless, about half of people with asthma in Europe are reported to be either partly controlled or uncontrolled[14,15]. Failure to control asthma symptoms may not only increase asthma exacerbations but also increase risk of other disease such as atrial fibrillation[16]. Few studies have investigated the possible associations between the levels of asthma symptom control and lung cancer risk.

The aim of the current study was to explore the potential associations of asthma overall as well as asthma status and symptom control with lung cancer incidence in a large prospective cohort study with a long follow-up duration.

Results

In total, 984 of the 62,791 participants developed lung cancer during a median follow-up of 21.1 years. Table 1 describes the distribution of baseline characteristics of participants by asthma categories. Compared to those without asthma, participants with non-active or active asthma (active asthma was defined if participants with asthma confirmed symptoms of wheezing or reported using asthma medication at the baseline survey) were more likely to be former or passive smokers, and to have allergic rhinitis, economic difficulties or a family history of cancer at baseline (Table 1). Similar patterns were found among participants with controlled or partially controlled asthma (the latter was defined as fulfilling one or more items based on the Global Initiative for Asthma guidelines) compared with participants without asthma (Supplementary Table S1). The proportion of controlled asthma was 37% among the participants with asthma and was 29% among the participants with active asthma.

There was no clear association between asthma overall and lung cancer incidence, with a HR of 1.19 (95% CI 0.91–1.57) after adjustment for smoking and other confounders (Table 2). Active asthma tended to be associated with an increased lung cancer incidence with an imprecise estimate (HR 1.29, 95% CI 0.95–1.75). Notably, partially controlled asthma showed an increased lung cancer incidence with an adjusted HR of 1.39 (95% CI 1.00–1.92). Neither non-active asthma nor controlled asthma was associated with the incidence of lung cancer (Table 2). In the adjusted model allergic rhinitis, as a possible confounder, was not associated with lung cancer incidence (HR 1.05, 95% CI 0.80–1.39). In addition, participants with both asthma and allergic rhinitis had a similar HR (1.25, 95% CI 0.86–1.82) for lung cancer incidence as participants with asthma and without allergic rhinitis (1.22, 95% CI 0.81–1.83) compared with those without asthma.

Results from the following sensitivity analyses provided supportive evidence for the above findings: (1) After excluding the first three-year follow-up, the associations of active asthma and partially controlled asthma with lung cancer incidence were strengthened (HR 1.41, 95% CI 1.04–1.93 for active asthma and HR 1.54, 95% CI 1.10–2.14 for partially controlled asthma, Table 3). After exclusion of the first five-year follow-up, it showed similar pattern of results as the originals (Supplementary Table S2). (2) The results after exclusion of asthma cases with a higher possibility of COPD using the post-bronchodilator fixed ratio of FEV_1/FVC or the lower limit of normal approach were similar to the original ones (Supplementary Tables S3 and S4). (3) Multiple imputation for missing data of all covariates including smoking showed comparable association estimates (Supplementary Table S5) compared with those before the imputation both in the primary cohort (Table 2) and in the cohort excluding the first 3-year follow-up (Table 3). (4) In the analysis using a negative control exposure (details are given in Supplementary Text, Figure and Table S6), migraine was inversely associated with heavy smoking in our study population. However, migraine was not associated with lung cancer incidence after adjustment for smoking, suggesting that our observed associations of active asthma and partially controlled asthma with lung cancer incidence were less likely biased by residual confounding due to smoking. (5) The results of competing

					Crude[1]		Adjusted[2]	
Asthma overall			**n/Cases**	**IR (per 1000 person-years)**	**HR**	**95% CI**	**HR**	**95% CI**
No			59,591/921	0.82	1.00	Reference	1.00	Reference
Yes			3200/63	1.08	1.29	1.00–1.67	1.19	0.91–1.57
	Asthma status							
		Non-active asthma	1090/14	0.69	1.01	0.59–1.71	0.94	0.55–1.61
		Active asthma	2110/49	1.28	1.41	1.06–1.88	1.29	0.95–1.75
	Asthma symptom control[3]							
		Controlled	1170/15	0.66	0.88	0.53–1.46	0.91	0.54–1.52
		Partially controlled	1622/42	1.47	1.58	1.16–2.16	1.39	1.00–1.92

Table 2. The associations of asthma overall, asthma status and levels of asthma symptom control with lung cancer incidence, the HUNT Study, 1995–97 to 2017 (n = 62,791). CI: confidence interval; HR: hazard ratio; IR: incidence rate. [1]Age was used as the time scale in the crude model. [2]Adjusted for sex, body mass index, smoking [(never, former (< 10, 10–20, and > 20 pack-years (pyrs)), current (< 10, 10–20, and > 20 pyrs)], passive smoking, alcohol consumption, physical activity, total sitting time daily, education, economic difficulties, family history of cancer and allergic rhinitis. Age was used as the time scale. *Tvc* option of the *stcox* command in Stata was used to model the non-proportional hazards for sex, smoking and economic difficulties in the adjusted models. [3]An "unknown" level of asthma symptom control is not shown due to limited lung cancer cases (n = 6).

					Crude[1]		Adjusted[2]	
Asthma overall			**n/Cases**	**IR (per 1000 person-years)**	**HR**	**95% CI**	**HR**	**95% CI**
No			58,211/834	0.88	1.00	Reference	1.00	Reference
Yes			3104/60	1.23	1.38	1.06–1.79	1.27	0.96–1.68
	Asthma status							
		Non-active asthma	1051/12	0.70	0.96	0.54–1.70	0.90	0.51–1.61
		Active asthma	2053/48	1.50	1.54	1.15–2.07	1.41	1.04–1.93
	Asthma symptom control[3]							
		Controlled	1152/15	0.78	0.96	0.57–1.60	0.98	0.58–1.66
		Partially controlled	1564/41	1.73	1.74	1.27–2.39	1.54	1.10–2.14

Table 3. The associations of asthma overall, asthma status and levels of asthma symptom control with lung cancer incidence after excluding the first three-year follow-up, the HUNT Study, 1995–97 to 2017 (n = 61,315). CI: Confidence interval; HR: Hazard ratio; IR: Incidence rate. [1]Age was used as the time scale in the crude model. [2]Adjusted for sex, body mass index, smoking [(never, former (< 10, 10–20, and > 20 pack-years (pyrs)), current (< 10, 10–20, and > 20 pyrs)], passive smoking, alcohol consumption, physical activity, total sitting time daily, education, economic difficulties, family history of cancer and allergic rhinitis. Age was used as the time scale. *Tvc* option of the *stcox* command in Stata was used to model the non-proportional hazards for sex, smoking and economic difficulties in the adjusted models. [3]An "unknown" level of asthma symptom control with limited lung cancer cases (n = 4) is not shown.

risk analysis excluding the influence of death showed a similar trend as our main results with wider CIs due to many cases of death (n = 15,653) (Supplementary Table S7).

Discussion

In this prospective cohort study with 984 incident lung cancer cases during a median follow-up of 21.1 years, we did not observe a clear association between asthma overall and lung cancer incidence. However, partially controlled asthma was associated with an increased lung cancer incidence. Adults with active asthma showed a tendency of increased lung cancer incidence. There was no association of non-active asthma or controlled asthma with the lung cancer incidence.

Previous meta-analysis studies have suggested a positive association between asthma and lung cancer[2,4,5]. We did not observe a clear association between asthma overall and lung cancer incidence. One of the explanations for this discrepancy may be due to residual confounding by smoking. Many of the studies included in the meta-analyses did not thoroughly address the role of smoking in the association[6–8]. The best way to address confounding by smoking is to study the association among never smokers or with certain lung cancer histologic types. Rosenberger et al. found that the positive association between asthma and lung cancer became weaker in a sub-analysis among never smokers or when lung cancer was restricted to adenocarcinoma, a histologic type

that is less strongly associated with smoking than other lung cancer subtypes[4]. We were not able to evaluate the associations in never smokers and with histologic types due to small number of lung cancer cases (e.g. there were only 56 lung cancer cases in the never smokers). Instead, we classified smoking status into detailed categories including pack-years in the adjusted model and performed sensitivity analyses such as multiple imputations and negative control exposure to address the possibility of residual confounding by smoking. Although we did not observe a clear association with asthma overall, we found that partially controlled asthma was associated with an increased incidence of lung cancer. Participants with active asthma had a tendency for increased lung cancer incidence. Analyses after multiple imputations for the missing data of all confounders including smoking showed similar pattern of results. The negative control exposure analysis using migraine also suggested that the observed associations were less likely biased by residual confounding from smoking, but residual confounding by smoking cannot completely be excluded. In line with our findings, Pirie and his colleagues included over half a million never smoking women from the UK Million Women Study who were followed up for more than 14 years and found that asthma requiring treatment was associated with an increased incidence of lung cancer[13].

Active airway inflammation linked with active asthma or partially controlled asthma may be associated with the lung carcinogenic process through elevated levels of free radicals and reduced levels of antioxidants[3], increased DNA damages and mutations[2], and permanent abnormality of the airways[17]. In our study 29% of the active asthma patients were controlled. As there was no association between controlled asthma and incidence of lung cancer, this would dilute the association so that we only observed a tentative association between active asthma and incidence of lung cancer. Thus, levels of asthma symptom control may reflect the levels of inflammation better than the definition of active asthma. On the other hand, participants with partially controlled asthma were likely to visit their physicians. This might have led to an increased referral to x-ray of thorax and thus a greater chance for screening of lung cancer. In addition, 50% of the participants with partially controlled asthma in our study reported having used inhaled corticosteroid (ICS) regularly. Previous studies have reported an independently inverse association of the use of ICS with lung cancer risk[18,19]. Our study showed similar HRs for lung cancer incidence among adults with partially controlled asthma who used ICS compared with those who did not use ICS (data not presented). As we did not have information on the dosage or the patients' compliance with ICS use, the potential influence of ICS on the risk of lung cancer warrants further investigation.

Our study is one of few prospective cohort studies that have investigated the potential associations of asthma overall, asthma status and symptom control with the incidence of lung cancer. We had a large and homogeneous study population and a long follow-up duration over 20 years. We also had information on a panel of potential confounders at baseline, which made it possible to minimize confounding. The information of lung cancer cases from the Cancer Registry of Norway is complete and accurate[20]. Misclassification of asthma due to early undiagnosed lung cancer, the so-called reverse causation, might not be an important problem in this study as we observed strengthened results after exclusion of the first three years of follow-up.

Our study has several limitations. First, misclassification of asthma was possible due to self-reporting. Nevertheless, self-reported asthma has been verified to be highly specific and reliable in many population studies[21]. In addition, the prevalence of asthma and active asthma (5.1% and 3.3% respectively) in our study was similar to a previous HUNT study using a slightly different definition of asthma[22] and another Nordic study[23]. In our study 37% of the participants with asthma were symptom controlled, which was comparable with the findings from other European studies[14,15]. All information on asthma was collected at baseline long before the diagnosis of lung cancer. Thus, the misclassification of asthma was likely to be non-differential that in general would lead to an underestimated association. However, asthma symptoms fluctuate over time. The association estimates generated from the one-time measure of asthma symptoms from HUNT2 may not reflect the true effect of the varying asthma symptoms on lung cancer incidence. Second, COPD may be misdiagnosed as asthma due to similar symptoms, especially among elderly people. This can bias the association between asthma and lung cancer. However, we excluded participants with possible COPD according to the GOLD definition at baseline and further excluded asthma cases with higher possibility of COPD in the sensitivity analyses. Third, we did not have information on air pollution, radon, or occupational exposure to asbestos and other carcinogenic agents, which are also important risk factors for lung cancer[24]. However, except for two smallest municipalities having had mining industries previously, there was nearly no industrial pollution in the northern area of Trøndelag during the time of HUNT2 study[22]. The level of indoor radon in the county was shown to be in the safety range (< 200 Bq/m3) in the national measurement during 1999–2000[25]. No asbestos-cement factories have existed in the county[26]. In any case, people with heavy exposure to asbestos should be minorities due to prohibition of importation and strict regulation to the use of asbestos in Norway since 1980[26]. At last, even if we have attempted to adjust for a large panel of potential confounders in our analyses, we cannot exclude the possibility of unknown confounding.

In conclusion, our study showed that participants with partially controlled asthma had an increased incidence of lung cancer. The finding suggested that proper control of asthma symptoms not only reduced asthma exacerbations but might also contribute to a reduced incidence of lung cancer.

Methods

Study design and population. The baseline data were derived from the second survey of The HUNT Study (HUNT2, 1995–1997). All adults aged 20 years or older living in the area of northern Trøndelag, Norway were invited to complete general questionnaires on health and lifestyle factors and undergo clinical examinations[27].

A total of 65,227 adults (69% of the invited) participated in HUNT2. Every participant was followed up from the date of participation in HUNT2 until the date of first diagnosis of lung cancer, the date of death or emigration from Norway or the end of follow-up on December 31, 2017, whichever came first. Lung cancer diagnoses were obtained from the Cancer Registry of Norway. Information on vital status and emigration was obtained from the National Population Registry.

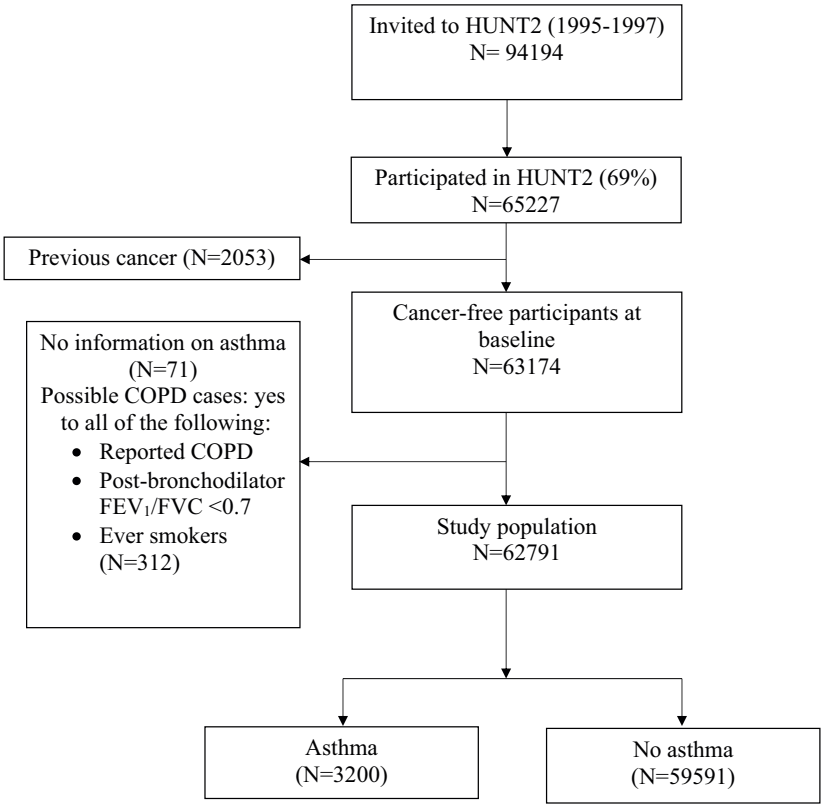


Figure 1. Flow chart of study participants.

We first excluded 2053 participants with previous cancer diagnoses before the baseline based on information from the Cancer Registry of Norway. Additionally, we excluded 71 participants without information on ever asthma. To minimize the influence by chronic obstructive pulmonary disease (COPD), we further excluded 312 adults who had possible COPD with all of the following criteria according to the Global Initiative for Chronic Obstructive Lung Disease (GOLD) definition[28]: reported doctor-diagnosed COPD, post-bronchodilator FEV_1/FVC < 0.7 and ever smokers at baseline. Reported doctor-diagnosed COPD was defined based on the question "Have you been diagnosed as having chronic bronchitis or emphysema by a doctor?". Lung function was measured by spirometry in HUNT2. The FEV_1/FVC ratio was calculated from forced expiratory volume in 1 s (FEV_1) and forced vital capacity (FVC)[28]. The post-bronchodilator fixed ratio (FEV_1/FVC < 0.7) is the recommended spirometric criterion for diagnosing COPD[28]. This left 62,791 participants in the primary cohort for analyses (Fig. 1).

Asthma definition. Detailed information on asthma history, symptoms and medication use was obtained from questionnaires[27]. Asthma was defined by affirmative answers to the following two questions: "Do you have, or have you had asthma?" in combination with "Have you been diagnosed as having asthma by a doctor?" (n = 3200). Asthma status was further categorized into active (n = 2110) and non-active (n = 1090) asthma. Participants were considered having active asthma if they confirmed symptoms of wheezing or reported using asthma medication in the last 12 months. The following four items were used to describe the level of asthma symptom control: 1) daytime symptoms more than twice weekly, 2) any night awakening, 3) need for reliever medications more than twice weekly or 4) any activity limitation based on the Global Initiative for Asthma (GINA) Global Strategy for Asthma Management and Prevention guidelines[29]. According to the GINA guidelines, controlled asthma refers to asthma without any of the above four items, partly controlled asthma refers to 1–2 items and uncontrolled asthma 3–4 items. We initially classified levels of asthma symptom control as controlled (n = 1170), partly controlled (n = 1227), uncontrolled (n = 395), and unknown (n = 408). Partly controlled and uncontrolled were collapsed into one category named "partially controlled" due to a small number of lung cancer cases in the uncontrolled category.

Other baseline variables. Weight and height were measured by health professionals at clinical examination. Body mass index (BMI) was calculated as weight in kilograms divided by height squared in meter (kg/m^2) and was grouped into three categories (< 25.0, 25.0–29.9, and ≥ 30.0 kg/m^2) according to the recommendations of the World Health Organization (WHO)[30]. Based on information of smoking status and pack-years, participants were classified into the detailed categories of smoking: never, former (< 10, 10–20, and > 20 pack-years (pyrs)) and current (< 10, 10–20, and > 20 pyrs). Other covariates were categorized as: passive smoking (never, only childhood, only adulthood, and both), alcohol consumption (never, 1–4, and ≥ 5 times/month), physical activity (inactive, low, moderate, and high), total sitting time daily (0–4, 5–7, and ≥ 8 h), education (< 10, 10–12, and ≥ 13 years), economic difficulty (yes/no) and family history of cancer (yes/no). Participants were considered having allergic rhinitis if reporting having allergic rhinitis in combination with use of allergy medication or allergic symptoms to pollen or pets. Missing information on each of the aforementioned variables was included in the analyses as an "unknown" category.

Ascertainment of lung cancer. By using the unique 11-digit personal identification number, participants' information from HUNT2 was linked to the Cancer Registry of Norway[31]. The International Classification of Diseases version 10 (ICD-10) codes used for registration of lung cancer are C33-C34[31]. Data from the Cancer Registry of Norway are reasonably accurate and complete[32].

Statistical analysis. Baseline characteristics of the participants were presented by asthma categories (no asthma, non-active asthma, and active asthma). We used Cox proportional hazard models to examine the potential associations of asthma overall, asthma status, and levels of asthma symptom control with lung cancer incidence. Crude and adjusted hazard ratios (HRs) with 95% confidence intervals (CIs) were calculated. Age was used as the underlying time variable. Potential confounders were selected based on previous knowledge[33–36] and directed acyclic graph (DAG). In the adjusted model, detailed categories of smoking status combined with pack-years was used to minimize confounding by smoking. The model also took account of sex, BMI, passive smoking, alcohol consumption, physical activity, total sitting time daily, education, economic difficulties, family history of cancer and allergic rhinitis. Allergic rhinitis was included in the model as a potential confounder because it commonly occurs together with asthma and has been reported to be inversely associated with lung cancer risk[34].

We assessed the proportional hazards assumption by Schoenfeld Residuals for exposures and all covariates. Apart from sex, smoking and economic difficulties, other covariates did not show evidence against proportional hazards assumption. We therefore used the *tvc* option of the *stcox* command in Stata to model the non-proportional hazards for sex, smoking and economic difficulties.

We performed several sensitivity analyses to test the robustness of our findings: (1) To address reverse causality by existing but undiagnosed lung cancer, we excluded both the first three-year and five-year follow-up. (2) To further minimize the misclassification of COPD as asthma, we used two ways to exclude asthma cases with a higher possibility of COPD: we excluded asthma cases who had smoked ten pack-years or more and were older than 40 years when getting the asthma diagnosis and had (a) post-bronchodilator $FEV_1/FVC < 0.7$ (fixed ratio criterion) (n = 104) or (b) post-bronchodilator FEV_1/FVC z score < -1.64 (lower limit of normal criterion) (n = 75). The fixed ratio criterion is the mostly used approach to define airflow limitation, whereas the lower limit of normal criterion overcomes the overestimation of the number of COPD among elderly[28]. (3) To address residual confounding by smoking and other covariates due to information missing, we conducted multivariable chained imputation with fully conditional specification (m = 10 imputed datasets) for the missing data of all covariates based on the assumption of missing at random. (4) To further address residual confounding by smoking, we performed analysis using migraine as a negative control exposure. This negative control exposure should be associated with the same confounder as the main exposure but not causally associated with the outcome [37]. The aim of analysis using a negative control exposure is to identify residual confounding that may have resulted in invalid causal inference for the main exposure-outcome association[37]. Previous study suggested that migraine was associated with smoking[38] but not with lung cancer. If we observed no association between migraine and lung cancer after adjustment for smoking, it indicated that the main exposure-outcome (asthma-lung cancer) association was less likely resulted from residual confounding by smoking. (5) To deal with possible competing risk due to death, a competing risk analysis based on Fine-Gray model was used[39]. All statistical analyses were performed with STATA/SE 15.1 (College Station, TX, USA).

Ethics approval

The study was approved by the Regional Committee for Medical and Health Research Ethics of South-East Norway (2015/78/REC South-East). All participants signed informed written consent on participation in HUNT. The study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

Data availability

Data from the HUNT Study is available on request to the HUNT Data Access Committee (hunt@ medisin.ntnu.no) when is used in research projects. The HUNT data access information describes the policy regarding data availability (https://www.ntnu.edu/hunt/data).

Received: 11 August 2020; Accepted: 11 February 2021
Published online: 25 February 2021

References

1. The Global Asthma Network. The Global Asthma Report 2018. (2018).
2. Santillan, A. A., Camargo, C. A. Jr. & Colditz, G. A. A meta-analysis of asthma and risk of lung cancer (United States). *Cancer Causes Control* **14**, 327–334 (2003).
3. Azad, N., Rojanasakul, Y. & Vallyathan, V. Inflammation and lung cancer: roles of reactive oxygen/nitrogen species. *J. Toxicol. Environ. Health. B Crit. Rev.* **11**, 1–15 (2008).
4. Rosenberger, A. *et al.* Asthma and lung cancer risk: a systematic investigation by the International Lung Cancer Consortium. *Carcinogenesis* **33**, 587–597 (2012).
5. Qu, Y. L. *et al.* Asthma and the risk of lung cancer: a meta-analysis. *Oncotarget* **8**, 11614–11620 (2017).
6. Boffetta, P., Ye, W., Boman, G. & Nyren, O. Lung cancer risk in a population-based cohort of patients hospitalized for asthma in Sweden. *Eur. Respir. J.* **19**, 127–133 (2002).
7. Hwang, C. Y. *et al.* Cancer risk in patients with allergic rhinitis, asthma and atopic dermatitis: a nationwide cohort study in Taiwan. *Int. J. Cancer* **130**, 1160–1167 (2012).
8. Ji, J. *et al.* Cancer risk in hospitalised asthma patients. *Br. J. Cancer* **100**, 829–833 (2009).
9. Brown, D. W., Young, K. E., Anda, R. F., Felitti, V. J. & Giles, W. H. Re: asthma and the risk of lung cancer. findings from the Adverse Childhood Experiences (ACE). *Cancer Causes Control* **17**, 349–350 (2006).
10. Colak, Y., Afzal, S., Nordestgaard, B. G. & Lange, P. Characteristics and Prognosis of Never–Smokers and smokers with asthma in the Copenhagen general population study. A Prospective Cohort Study. *Am. J. Respir. Crit. Care Med.* **192**, 172–181 (2015).
11. Mills, P. K., Beeson, W. L., Fraser, G. E. & Phillips, R. L. Allergy and cancer: organ site-specific results from the Adventist Health Study. *Am. J. Epidemiol.* **136**, 287–295 (1992).
12. Talbot-Smith, A., Fritschi, L., Divitini, M. L., Mallon, D. F. & Knuiman, M. W. Allergy, atopy, and cancer: a prospective study of the 1981 Busselton cohort. *Am. J. Epidemiol.* **157**, 606–612 (2003).
13. Pirie, K., Peto, R., Green, J., Reeves, G. K. & Beral, V. Lung cancer in never smokers in the UK Million Women Study. *Int. J. Cancer* **139**, 347–354 (2016).
14. Cazzoletti, L. *et al.* Asthma control in Europe: a real-world evaluation based on an international population-based study. *J. Allergy Clin. Immunol.* **120**, 1360–1367 (2007).
15. Demoly, P. *et al.* Prevalence of asthma control among adults in France, Germany, Italy, Spain and the UK. *Eur. Respir. Rev.* **18**, 105–112 (2009).
16. Cepelis, A. *et al.* Associations of asthma and asthma control with atrial fibrillation risk: results from the Nord–Trøndelag health study (HUNT). *JAMA Cardiol.* **3**, 721–728 (2018).
17. Roche, W. R. Inflammatory and structural changes in the small airways in bronchial asthma. *Am. J. Respir. Crit. Care Med.* **157**, S191-194 (1998).
18. Parimon, T. *et al.* Inhaled corticosteroids and risk of lung cancer among patients with chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **175**, 712–719 (2007).
19. Lee, C. H. *et al.* Inhaled corticosteroid use and risks of lung cancer and laryngeal cancer. *Respir. Med.* **107**, 1222–1233 (2013).
20. Cancer Registry of Norway. (2017). http://www.kreftregisteret.no/.
21. Toren, K., Brisman, J. & Jarvholm, B. Asthma and asthma-like symptoms in adults assessed by questionnaires. A literature review. *Chest* **104**, 600–608 (1993).
22. Langhammer, A., Johnsen, R., Holmen, J., Gulsvik, A. & Bjermer, L. Cigarette smoking gives more respiratory symptoms among women than among men The Nord-Trøndelag Health Study (HUNT). *J. Epidemiol. Commun. Health* **54**, 917–922 (2000).
23. Dahl, R. & Bjermer, L. Nordic consensus report on asthma management. *Respir. Med.* **94**, 299–327 (2000).
24. Subramanian, J. & Govindan, R. Lung cancer in never smokers: a review. *J. Clin. Oncol.* **25**, 561–570 (2007).
25. Kommunekartlegginger-radon (Municipal Surveys-Radon). (2013). http://www.nrpa.no/temaartikler/90379/kart-og-data-radon .
26. Ulvestad, B., Kjaerheim, K., Møller, B. & Andersen, A. Incidence trends of mesothelioma in Norway, 1965–1999. *Int. J. Cancer* **107**, 94–98 (2003).
27. Krokstad, S. *et al.* Cohort profile: the HUNT study, Norway. *Int. J. Epidemiol.* **42**, 968–977 (2013).
28. Vogelmeier, C. F. *et al.* Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report. GOLD Executive Summary. *Am. J. Respir. Crit. Care Med.* **195**, 557–582 (2017).
29. 2019 GINA Report. Global Strategy for Asthma Management and Prevention 82019 update. (2019).
30. WHO. *Preventing and Managing the Global Epidemic [Report of a WHO Consultation on Obesity]* (World Health Organization, Geneva, 2004).
31. Cancer Registry of Norway. *Cancer in Norway 2014-Cancer Incidence, Mortality, Survival and Prevalence in Norway* (Cancer Registry of Norway, Oslo, 2015).
32. Larsen, I. K. *et al.* Data quality at the Cancer Registry of Norway: an overview of comparability, completeness, validity and timeliness. *Eur. J. Cancer* **45**, 1218–1231 (2009).
33. Jiang, L. *et al.* Prolonged sitting, its combination with physical inactivity and incidence of lung cancer: prospective data from the HUNT Study. *Front. Oncol.* **9**, 101 (2019).
34. Kantor, E. D., Hsu, M., Du, M. & Signorello, L. B. Allergies and asthma in relation to cancer risk. *Cancer Epidemiol. Biomarkers Prev.* **28**, 1395–1403 (2019).
35. Turner, M. C. *et al.* Cancer mortality among US men and women with asthma and hay fever. *Am. J. Epidemiol.* **162**, 212–221 (2005).
36. Sun, Y. Q. *et al.* Associations of serum 25-hydroxyvitamin D level with incidence of lung cancer and histologic types in Norwegian adults: a case-cohort analysis of the HUNT study. *Eur. J. Epidemiol.* **33**, 67–77 (2018).
37. Lipsitch, M., Tchetgen, E. T. & Cohen, T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology (Cambridge, Mass.)* **21**, 383 (2010).
38. López-Mesonero, L. *et al.* Smoking as a precipitating factor for migraine: a survey in medical students. *J. Headache Pain* **10**, 101 (2009).
39. Austin, P. & Fine, J. Practical recommendations for reporting Fine–Gray model analyses for competing risk data. *Stat. Med.* **36**, 4391–4400 (2017).

Acknowledgments

The HUNT Study is collaboration between HUNT Research Centre (Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology—NTNU), Nord-Trøndelag County Council, Central Norway Regional Health Authority and the Norwegian Institute of Public Health.

Author contributions

L.J. conducted statistical analyses, interpreted results and wrote the initial draft of the manuscript. Y.Q.S. and X.M.M. contributed to the study design and statistical analyses. A.L. and X.M.M. were responsible for data

collection. All authors participated in the data interpretation, contributed to the manuscript writing with important intellectual content and approved the final version of the manuscript.

Funding
LJ was supported by funding from the Stiftelsen Dam through the Norwegian Cancer Society (project no. 90539400) and top-up funding from the collaboration partner between the Liaison Committee for Education, Research and Innovation in Central Norway and Central Norway Regional Health Authority. YQS was supported by a researcher grant from The Liaison Committee for Education, Research and Innovation in Central Norway. BMB works in a research unit funded by Stiftelsen Kristian Gerhard Jebsen; Faculty of Medicine and Health Sciences, NTNU; The Liaison Committee for Education, Research and Innovation in Central Norway; and Medical Research Council Integrative Epidemiology Unit at the University of Bristol. None of the funding sources was involved in any aspect of the study design, conduct, analysis, interpretation of data, or writing the report.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-84012-3.

Correspondence and requests for materials should be addressed to L.J.

Reprints and permissions information is available at www.nature.com/reprints.

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

Supplementary material: Paper II

Supplementary Table S1. Distribution of baseline characteristics according to levels of asthma symptom control in the HUNT2 Study, 1995-1997 (N = 62,791).

Variables	No asthma	Controlled asthma	Partially controlled asthma	Unknown
Number of subjects	**59,591**	**1170**	**1622**	**408**
Age (years)	49.5±17.0	47.0±16.2	52.1±17.5	48.5±18.8
Body mass index (kg/m^2)	26.3±4.0	26.7±4.3	27.4±5.0	27.3±5.2
Number of lung cancer cases (%)	**921(1.6)**	**15(1.3)**	**42 (2.6)**	**6 (1.5)**
Sex, % (women/men)	53.0/47.0	53.4/46.6	55.0/45.0	55.6/44.4
Allergic rhinitis, % (no/yes/unknown)	70.5/5.5/24.0	41.1/45.2/13.7	34.5/51.2/14.4	36.5/47.8/15.7
Smoking status, % (never/current/former/unknown)	43.0/28.6/26.3/2.2	41.1/27.4/29.8/1.7	36.3/28.2/33.2/2.3	39.2/32.3/25.5/2.9
Passive smoking, % (never/ever/unknown)	18.5/79.6/1.9	16.9/81.8/1.3	14.4/83.5/2.0	12.8/85.5/1.7
Alcohol consumption (times/month), % (never/≥1/unknown)	34.6/56.9/8.6	31.9/60.5/7.6	40.5/51.1/8.5	39.7/51.7/8.6
Physical activity, % (inactive[1]/active[2]/unknown)	21.6/48.0/30.4	21.9/50.5/27.6	23.8/44.3/31.9	26.5/50.7/22.8
Total sitting time daily (hours), % (<8/≥8/unknown)	48.1/27.8/24.1	54.7/33.0/12.3	53.8/32.7/13.5	50.7/34.3/15.0
Education (years), % (<10/≥10/unknown)	34.0/61.0/5.1	30.7/65.4/3.9	40.2/53.1/6.7	36.5/56.9/6.6
Economic difficulties, % (no/yes/unknown)	48.0/21.1/30.9	56.8/26.1/17.2	45.3/31.8/23.0	40.0/34.8/25.3
Family history of cancer, % (no/yes)	75.0/25.0	70.8/29.2	68.6/31.4	76.5/23.5

HUNT: Nord-Trøndelag Health Study.

Data are given as mean ± standard deviation or percentage of subjects in each asthma category.

[1] Inactive: no physical activity or only light physical activity ≤2 h per week.

[2] Active: physical activity level from low to high.

Supplementary Table S2. The associations of asthma overall, asthma status and levels of asthma symptom control with lung cancer incidence after excluding the first five-year follow-up, the HUNT Study, 1995-97 to 2017 (N = 59,944)

					Crude[1]		Adjusted[2]	
Asthma overall			n/Cases	IR (per 1000 person-years)	HR	95% CI	HR	95% CI
No			56,941/784	0.94	1.00	Reference	1.00	Reference
Yes			3003/54	1.26	1.33	1.01-1.75	1.22	0.91-1.64
	Asthma status							
		Non-active asthma	1020/12	0.80	1.02	0.58-1.81	0.95	0.53-1.70
		Active asthma	1983/42	1.51	1.46	1.07-1.99	1.33	0.96-1.85
	Asthma symptom control[3]							
		Controlled	1133/15	0.89	1.01	0.61-1.68	1.02	0.60-1.72
		Partially controlled	1495/35	1.69	1.61	1.15-2.26	1.43	1.00-2.04

CI: Confidence interval; HR: Hazard ratio; IR: Incidence rate.

[1] Age was used as the time scale in the crude model.

[2] Adjusted for sex, body mass index, smoking, passive smoking, alcohol consumption, physical activity, total sitting time daily, education, economic difficulties, family history of cancer, and allergic rhinitis. Age was used as the time scale. *Tvc* option of the *stcox* command in Stata was used to model the non-proportional hazards for sex, smoking, and economic difficulties in the adjusted models.

[3] An "unknown" level of asthma symptom control with limited lung cancer cases ($n = 4$) is not shown.

Supplementary Table S3. The associations of asthma overall, asthma status and levels of asthma symptom control with lung cancer incidence after excluding participants with asthma who also had post bronchodilator FEV_1/FVC <0.7, smoking package years ≥10.0 and age at asthma diagnosis >40 years, the HUNT Study, 1995-97 to 2017 (N = 62,687)

					Crude[1]		Adjusted[2]	
Asthma overall			n/Cases	IR (per 1000 person-years)	HR	95% CI	HR	95% CI
No			59,591/921	0.82	1.00	Reference	1.00	Reference
Yes			3096/55	0.96	1.19	0.90-1.56	1.20	0.90-1.60
	Asthma status							
		Non-active asthma	1085/14	0.69	1.01	0.60-1.72	0.97	0.57-1.66
		Active asthma	2011/41	1.11	1.26	0.92-1.72	1.31	0.94-1.82
	Asthma symptom control[2]							
		Controlled	1148/13	0.58	0.78	0.45-1.35	0.86	0.49-1.50
		Partially controlled	1540/36	1.31	1.46	1.04-2.03	1.46	1.03-2.07

CI: confidence interval; HR: hazard ratio; IR: incidence rate.

[1] Age was used as the time scale in the crude model.

[2] Adjusted for sex, body mass index, smoking, passive smoking, alcohol consumption, physical activity, total sitting time daily, education, economic difficulties, family history of cancer, and allergic rhinitis. Age was used as the time scale. *Tvc* option of the *stcox* command in Stata was used to model the non-proportional hazards for sex, smoking, and economic difficulties in the adjusted model.

[3] An "unknown" level of asthma symptom control is not shown due to limited lung cancer cases ($n = 6$).

Supplementary Table S4. The associations of asthma overall, asthma status and levels of asthma symptom control with lung cancer incidence after excluding participants with asthma who also had post bronchodilator FEV_1/FVC z score <-1.64, smoking package years ≥10.0 and age at asthma diagnosis >40 years, the HUNT Study, 1995-97 to 2017 (N = 62,716)

					Crude[1]		Adjusted[2]	
Asthma overall			n/Cases	IR (per 1000 person-years)	HR	95% CI	HR	95% CI
No			59,591/921	0.82	1.00	Reference	1.00	Reference
Yes			3125/58	1.01	1.23	0.94-1.61	1.21	0.91-1.60
	Asthma status							
		Non-active asthma	1087/14	0.69	1.01	0.60-1.71	0.96	0.56-1.64
		Active asthma	2038/44	1.18	1.32	0.98-1.79	1.32	0.96-1.82
	Asthma symptom control[2]							
		Controlled	1157/13	0.58	0.77	0.45-1.34	0.82	0.47-1.44
		Partially controlled	1560/39	1.40	1.55	1.12-2.13	1.49	1.06-2.08

CI: confidence interval; HR: hazard ratio; IR: incidence rate.

[1] Age was used as the time scale in the crude model.

[2] Adjusted for sex, body mass index, smoking, passive smoking, alcohol consumption, physical activity, total sitting time daily, education, economic difficulties, family history of cancer, and allergic rhinitis. Age was used as the time scale. *Tvc* option of the *stcox* command in Stata was used to model the non-proportional hazards for sex, smoking, and economic difficulties in the adjusted model.

[3] An "unknown" level of asthma symptom control is not shown due to limited lung cancer cases ($n = 6$).

Supplementary Table S5. The associations of asthma overall, asthma status and levels of asthma symptom control with lung cancer incidence after multiple imputations, the HUNT Study, 1995-97 to 2017 in the primary analysis cohort and in cohort after excluding the first three-year follow-up

			Primary cohort (N = 62,791)		Cohort after excluding the first 3-year follow-up (N = 61,315)	
Asthma overall			Adjusted[1] HR	95% CI	Adjusted[1] HR	95% CI
No			1.00	Reference	1.00	Reference
Yes			1.14	0.85-1.52	1.21	0.91-1.62
	Asthma status					
		Non-active asthma	0.90	0.52-1.56	0.86	0.48-1.55
		Active asthma	1.23	0.90-1.70	1.35	0.98-1.86
	Asthma symptom control[2]					
		Controlled	0.86	0.51-1.47	0.95	0.56-1.60
		Partially controlled	1.33	0.95-1.87	1.46	1.04-2.06

CI: confidence interval; HR: hazard ratio; IR: incidence rate.

[1] Adjusted for sex, body mass index, smoking, passive smoking, alcohol consumption, physical activity, total sitting time daily, education, economic difficulties, family history of cancer, and allergic rhinitis. Age was used as the time scale. *Tvc* option of the *stcox* command in Stata was used to model the non-proportional hazards for sex, smoking, and economic difficulties in the adjusted models.

[2] An "unknown" level of asthma symptom control is not shown due to limited lung cancer cases ($n = 6$).

Supplementary Text. Analysis using negative control exposure

The aim of analysis using a negative control exposure is to identify residual confounding that may have resulted in invalid causal inference for the main exposure-outcome association[1]. In the current study, we used "migraine" as the negative control exposure to detect residual confounding by smoking in the observed asthma-lung cancer association (Supplementary Figure). "Migraine" was chosen as the negative control exposure was because it is associated with the confounder (smoking)[2,3], but not causally associated with the outcome (lung cancer). We expected to observe a null association between migraine and lung cancer after adjustment for smoking, suggesting that the observed asthma-lung cancer association was less likely biased by the residual confounding of smoking.

Based on our study population (N = 62,791), we excluded participants without information on headache. This left 49,945 participants to study the relationship between migraine and lung cancer incidence. Participants with migraine were those who answered yes to the question "Have you suffered from headache during the last 12 months?" and specified the type of headache as "migraine", and the rest were regarded as no migraine. We adjusted for smoking status in model 1. In model 2, we adjusted for the same confounders as in our primary study (Supplementary Table S6).

The negative control exposure analysis showed that fewer participants with migraine were heavy smokers (> 20.1 pack-years) than participants without migraine (6.1% vs 10.3%). There was an inverse association between migraine and lung cancer incidence without adjustment for smoking (crude HR 0.61, 95% CI 0.43-0.86). After adjustment for smoking the association between migraine and lung cancer became less clear (HR 0.75, 95% CI 0.53-1.06), and additional adjustment for the same confounders as in the primary study did not have material changes in the result. This indicated that our observed associations of active asthma and partially controlled asthma with increased lung cancer incidence were less likely biased by residual confounding due to smoking.

Supplementary Figure. DAG for asthma (as the main exposure), migraine (as the negative control exposure) and incidence of lung cancer (as the outcome)

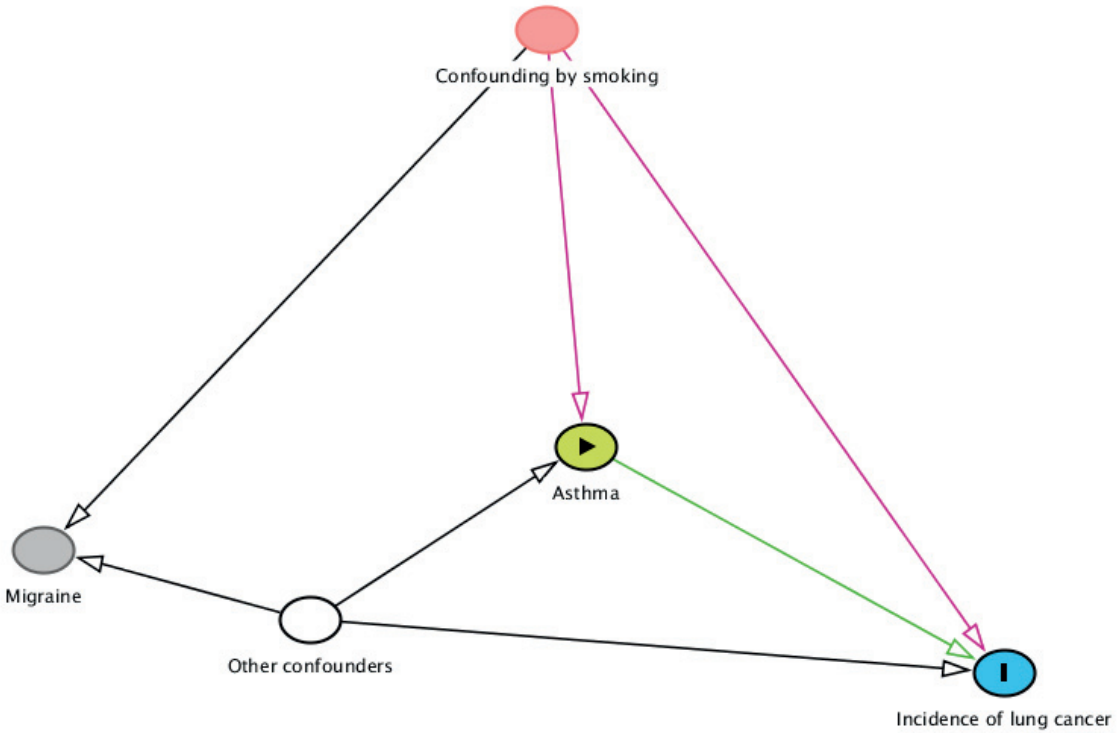


Figure legend. This DAG was created by using DAGitty V 3.0[4] (http://www.dagitty.net/dags.html).

Supplementary Table S6. Negative control using migraine as an alternative exposure to address residual confounding by smoking for the association of asthma with lung cancer incidence, the HUNT Study, 1995-97 to 2017 (N = 49,945)

				Crude[1]		Adjusted[2]		Adjusted[3]	
	n	Cases	IR (per 1000 person-years)	HR	95% CI	HR	95% CI	HR	95% CI
Non-migraine	45,750	721	0.83	1.00	Reference	1.00	Reference	1.00	Reference
Migraine	4195	35	0.41	0.61	0.43-0.86	0.75	0.53-1.06	0.75	0.53-1.05

CI: confidence interval; HR: hazard ratio; IR: incidence rate.

[1] Age was used as the time scale in the crude model.

[2] Adjusted for smoking [(never, former (≤10.0, 10.1-20.0, and >20.1 pack-years (pyrs)), current (≤10.0, 10.1-20.0, and >20.1 pyrs)]. Age was used as the time scale.

[3] The same confounders as for the association between asthma and lung cancer incidence: sex, body mass index, smoking, passive smoking, alcohol consumption, physical activity, total sitting time daily, education, economic difficulties, family history of cancer, and allergic rhinitis. Age was used as the time scale.

Supplementary Table S7. The associations of asthma overall, asthma status and levels of asthma symptom control with lung cancer incidence taking into account competing risk due to death, the HUNT Study, 1995-97 to 2017 (N = 62,791)

Asthma overall			n/Cases	IR (per 1000 person-years)	Adjusted[1] SHR	95% CI
No			59,591/921	0.82	1.00	Reference
Yes			3200/63	1.08	1.10	0.84-1.44
	Asthma status					
		Non-active asthma	1090/14	0.69	0.86	0.50-1.47
		Active asthma	2110/49	1.28	1.20	0.88-1.62
	Asthma symptom control[2]					
		Controlled	1170/15	0.66	0.86	0.51-1.46
		Partially controlled	1622/42	1.47	1.26	0.92-1.74

CI: confidence interval; IR: incidence rate; SHR: sub-distribution hazard ratio.

[1] Adjusted for sex, body mass index, smoking, passive smoking, alcohol consumption, physical activity, total sitting time daily, education, economic difficulties, family history of cancer, and allergic rhinitis. Age was used as the time scale. *Tvc* option of the *stcox* command in Stata was used to model the non-proportional hazards for sex, smoking, and economic difficulties in the adjusted model.

[2] An "unknown" level of asthma symptom control is not shown due to limited lung cancer cases ($n = 6$).

References

1.Lipsitch, M., Tchetgen, E.T. & Cohen, T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology (Cambridge, Mass.)* **21**, 383 (2010).
2.Taylor, F.R. Tobacco, Nicotine, and Headache. *Headache* **55**, 1028-1044 (2015).
3.Aamodt, A., Stovner, L., Hagen, K., Bråthen, G. & Zwart, J. Headache prevalence related to smoking and alcohol use. The Head-HUNT Study. *Eur. J. Neurol.* **13**, 1233-1238 (2006).
4.Textor, J., van der Zander, B., Gilthorpe, M.S., Liśkiewicz, M. & Ellison, G.T. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *Int. J. Epidemiol.* **45**, 1887-1894 (2016).

This article is under review for publication and is therefore not included.

Paper III

ISBN 978-82-326-5898-5 (printed ver.)
ISBN 978-82-326-5415-4 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU
Norwegian University of
Science and Technology