UNIVERSITY OF COPENHAGEN
FACULTY OF HEALTH AND MEDICAL SCIENCES

Doctoral thesis

From early undiagnosed to clinical chronic obstructive pulmonary disease

Yunus Çolak, MD, PhD

This thesis has been submitted to the Graduate School of Health and Medical Sciences, University of Copenhagen, March 2021

Doctoral thesis

From early undiagnosed to clinical chronic obstructive pulmonary disease

Yunus Çolak, MD, PhD

Contents

1. Çolak Y, Afzal S, Nordestgaard BG, Vestbo J, Lange P. Prognosis of asymptomatic and symptomatic, undiagnosed COPD in the general population in Denmark: a prospective cohort study. *Lancet Respir Med*. 2017; 5(5): 426-434.

2. Çolak Y, Afzal S, Nordestgaard BG, Vestbo J, Lange P. Young and middle-aged adults with airflow limitation according to lower limit of normal but not fixed ratio have high morbidity and poor survival: a population-based prospective cohort study. *Eur Respir J*. 2018; 51(3): 1702681.

3. Çolak Y, Nordestgaard BG, Vestbo J, Lange P, Afzal S. Prognostic significance of chronic respiratory symptoms in individuals with normal spirometry. *Eur Respir J*. 2019; 54(3): 1900734.

4. Çolak Y, Afzal, Nordestgaard BG, Vestbo J, Lange P. Prevalence, characteristics, and prognosis of early chronic obstructive pulmonary disease. The Copenhagen General Population Study. *Am J Respir J Crit Care Med*. 2020; 201(6): 671-680.

5. Çolak Y, Nordestgaard BG, Vestbo J, Lange P, Afzal S. Comparison of five major airflow limitation criteria to identify high-risk individuals with COPD: a contemporary population-based cohort. *Thorax*. 2020;75(11): 944-954.

6. Çolak Y, Afzal S, Nordestgaard BG, Lange P, Vestbo J. Importance of early COPD in young adults for development of clinical COPD: Findings from the Copenhagen General Population Study. *Am J Respir J Crit Care Med*. 2020; DOI:10.1164/rccm.202003-0532OC.

Preface

The doctoral thesis is based on research conducted during my stay at the Department of Clinical Biochemistry and Department of Internal Medicine at Herlev and Gentofte Hospital, Copenhagen University Hospital, from September 2016 until December 2020.

Firstly, I have always believed that helping others is my purpose in life and have therefore chosen this path as a medical doctor and a researcher. Thus, I am very grateful to Allah (c.c.) for been given endless opportunities in life. All the credit is due to Him and only the mistakes have been mine.

Secondly, I would like to express my gratitude towards all my mentors during all these years, Peter Lange, Børge G. Nordestgaard, Shoaib Afzal, and Jørgen Vestbo. Towards Peter and Jørgen for inspiring me to pursue the field of Respiratory Medicine for more than 10 years ago, for encouraging me and providing endless opportunities to advance in the field of obstructive lung diseases, and for being role models to strive for as a medical doctor and a researcher. Towards Børge and Shoaib for teaching me the key elements of good research, including thoroughness, extensive self-discipline, hard work, and the constant struggle for perfectionism, and for encouraging me and providing endless opportunities to advance in the field of epidemiology. I would also like to express my gratitude towards the steering committees of the Copenhagen General Population Study and the Copenhagen City Heart Study for conceiving such unique studies and allowing me to work with them, and the staff belonging to these studies for their tireless efforts in bringing these ideas to life.

Lastly, I wish to express my gratitude towards my parents, Adnan and Navruz, for their constant love and support and for teaching me the most important things in life. Without you, I would not be

who I am. I also wish to express my gratitude towards my little sister, Esma Nur, for her constant love and support.

Not least, I will forever be grateful to my dear wife, Sultan, without her unyielding love and support and her endless faith in me, this would not be possible. Thank you for being my companion and best friend through life.

For my beloved children Sâlih and Sevde

Summary in English

This doctoral thesis is formed by six original articles conducted during my stay at the Department of Clinical Biochemistry and Department of Internal Medicine at Herlev and Gentofte Hospital, Copenhagen University Hospital. The doctoral thesis consists of three parts:

- **Part I** is a review of emerging evidence regarding origins of chronic obstructive pulmonary disease (COPD) from early undiagnosed to clinical disease in the context of the main findings from the six original articles.
- **Part II** is a description of the applied methods in the six original articles with a critical review.
- **Part III** comprises the six original articles published in the *Lancet Respiratory Medicine*, *European Respiratory Journal*, *American Journal of Respiratory and Critical Care Medicine*, and *Thorax*.

COPD is associated with increased morbidity and mortality. A potential explanation for the increased burden may be delayed diagnosis and treatment in the disease course when irreversible lung damage has already taken place. However, COPD seems to have its origin in early life and develops gradually over many years. Individuals that will develop COPD may be identified early before disease onset allowing for implementation of preventive measures to halt progression. Indeed, individuals identified with early airflow limitation or chronic respiratory symptoms display an increased risk of hospitalisations due to obstructive lung disease and early death. Importantly, up to one-fourth of younger individuals identified with early COPD seems to develop clinical COPD 10 years later. Although there may be potential overdiagnosis and overtreatment, this group still displays impaired respiratory health and may benefit from preventive measures.

Summary in Danish

Denne doktorafhandling er dannet ud fra seks originale artikler gennemført under mit ophold på Klinisk Biokemisk Afdeling og Medicinsk Afdeling på Herlev og Gentofte Hospital. Doktorafhandlingen består af tre dele:

- **Del I** er en gennemgang af evidensen vedrørende oprindelsen af kronisk obstruktiv lungesygdom (KOL) fra tidligt udiagnosticeret til klinisk sygdom set i lyset af de vigtigste fund fra de seks originale artikler.
- **Del II** er en beskrivelse af de anvendte metoder i de seks originale artikler med en kritisk gennemgang.
- **Del III** omfatter de seks originale artikler offentliggjort i *Lancet Respiratory Medicine*, *European Respiratory Journal*, *American Journal of Respiratory and Critical Care Medicine* og *Thorax*.

KOL er associeret med øget sygelighed og dødelighed. En mulig forklaring på dette kan være forsinket diagnose og behandling i sygdomsforløbet, når irreversibel lungeskade allerede har fundet sted. KOL synes imidlertid at have sin oprindelse i det tidlige liv og udvikler sig gradvist over mange år. Individer, der udvikler KOL, kan muligvis identificeres tidligt inden sygdomsudbrud, hvor relevant forebyggende tiltag kan implementeres for at bremse progression. Individer identificeret med tidlig luftvejsobstruktion eller kroniske luftvejssymptomer har en øget risiko for indlæggelser på grund af obstruktiv lungesygdom og tidlig død. Op til en fjerdedel af yngre personer, der er identificeret med tidlig KOL, synes at udvikle klinisk KOL 10 år senere. Selvom der kan være potentiel overdiagnosticering og overbehandling, har denne gruppe påvirket lungehelbred og kan muligvis få gavn af forebyggende tiltag.

Part I – From early undiagnosed to clinical chronic obstructive pulmonary disease

Introduction

Worldwide, chronic obstructive pulmonary disease (COPD) is a prevalent disease in middle-aged and old adults and one of the leading causes of morbidity and mortality, a scenario likely to persist for many years.[1] A potential explanation for the increased burden of disease is that patients with COPD are often diagnosed and begin treatment very late in the disease course, that is, when severe airflow limitation has already developed.

It is now increasingly evident that COPD has its origin in early life and develops gradually over many years.[2,3] Susceptible individuals that will develop COPD later in life could therefore possibly be identified at an earlier age before disease onset allowing for implementation of preventive measures thereby halting progression.[4] Such an initiative may not only lead to a lower incidence and prevalence but also to milder forms of COPD with higher treatment potential. However, COPD has always been considered a disease of the elderly, and its presence in younger adults has not been investigated thoroughly. Furthermore, over the last decades, focus has mainly been on older patients with established severe COPD, as these comprise a significant proportion of the clinical consultations.[5]

Like other diseases such as diabetes, it was once believed that a prodromal phase of COPD exists, which could be used to identify high-risk individuals mainly among smokers.[6] Incomplete evidence that these high-risk individuals necessarily progress to clinical COPD, at some stage, however, ended the idea of early prevention and intervention.[7,8] Nonetheless, the idea was resurrected when an international group of experts recently proposed an operational definition for early COPD by distinguishing “early” from “mild” disease in order to facilitate more research in the field.[4] In this review following such developments, I will focus on COPD from early undiagnosed to clinical disease.

Natural history of airflow limitation and COPD

COPD is characterised by chronic airflow limitation that is believed to arise from a combination of airway and alveolar abnormalities (often denoted as small airways disease and pulmonary emphysema); however, the relative contribution of these two components may vary individually from patient to patient.[9] Development of airflow limitation, identified as a low ratio of forced expiratory volume in 1 second (FEV_1) to forced vital capacity (FVC), is usually preceded by long-term exposure to noxious particles or gases such as tobacco smoke, but host factors and gene-environment interactions are now becoming increasingly recognised to have a significant influence as well.[10]

Normal lung development is characterised by a rapid increase in lung function from birth during childhood with peak in adulthood around age 20-25 years, followed by a plateau phase with preservation of maximally attained lung function for approximately 5-10 years before a steady normal age-related decline occurs.[3] Since the seminal study by Charles Fletcher and colleagues in the 1970s, development and progression of COPD has been thought to be strongly linked to an acceleration of the normal age-related decline of lung function often due to active smoking.[11,12] Around the same time, another study by Benjamin Burrows and colleagues already suggested that COPD may have its origin in early life,[13] but this received less attention.

Over the years, the prevailing paradigm of COPD pathogenesis has been accelerated lung function decline in susceptible individuals, mainly among smokers.[14] Not until 40 years later, Peter Lange and colleagues challenged this paradigm by demonstrating that only half of patients with COPD developed the disease due to accelerated lung function decline during adult life, whereas the other half developed it due to low maximally attained lung function in early adulthood, suggesting that accelerated lung function decline is not an obligate feature of COPD.[15] Thus, the existence of

several lung function trajectories that may lead to development of COPD is now widely accepted, but only two have been identified so far (Figure 1).[9,16] Exposure to risk factors during different stages of life may affect development and preservation of normal lung function from birth until old age, perhaps already prenatally, and increase risk of airflow limitation and COPD.[17]

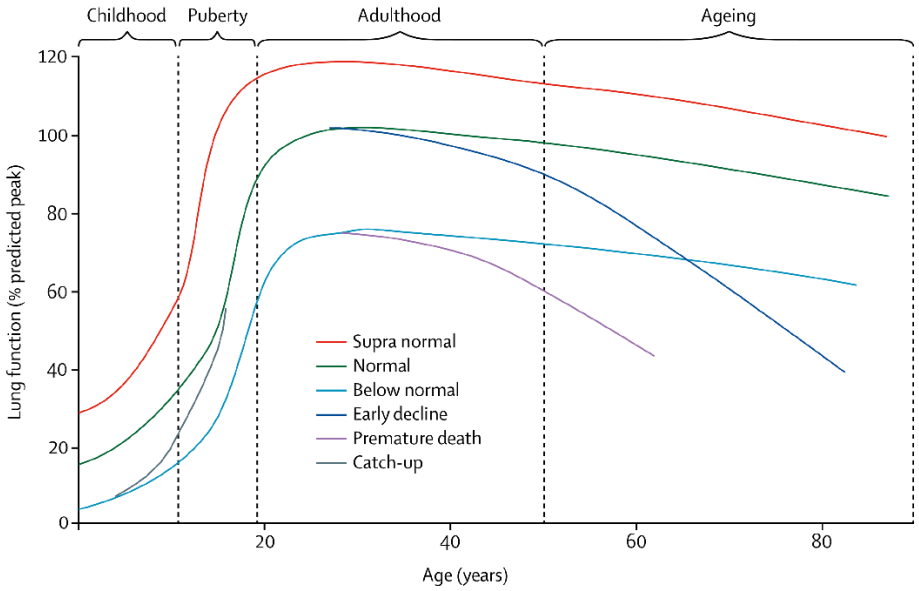

Figure 1 | Potential lung function trajectories from birth to death that may lead to development of COPD. Adapted from Agustí & Faner.[16]

Current clinical practice

Most countries follow the Global Initiative for Chronic Obstructive Lung Disease (GOLD) recommendations for the diagnosis and management of COPD in clinical practice.[10] Accordingly, COPD should be considered in all patients aged ≥40 years complaining of respiratory symptoms with a history of exposure to relevant risk factors such as tobacco smoke. A spirometry confirms the diagnosis by demonstrating presence of airflow limitation defined as a fixed ratio of post-bronchodilator FEV_1/FVC <0.70. Management recommendations are based on disease severity defined by degree of airflow limitation, symptom burden, and history of exacerbation (an acute

worsening of symptoms leading to additional therapy). While degree of airflow limitation is graded according to FEV_1 as % of predicted value, symptom burden is evaluated using modified Medical Research Council Dyspnoea Scale (mMRC) and/or COPD assessment test (CAT). A patient with COPD is classified according to degree of airflow limitation (GOLD 1-4) and symptom burden and exacerbation history (GOLD A-D) (Figure 2). The main treatment strategy is to reduce symptoms and prevent frequency and severity of future exacerbations, which mostly involves different combination therapies of long-acting beta-2-agonist (LABA), long-acting muscarinic antagonist (LAMA), and inhaled corticosteroid (ICS). Smoking cessation or intervention against other relevant exposures responsible for development of COPD should also be regarded as part of the treatment.

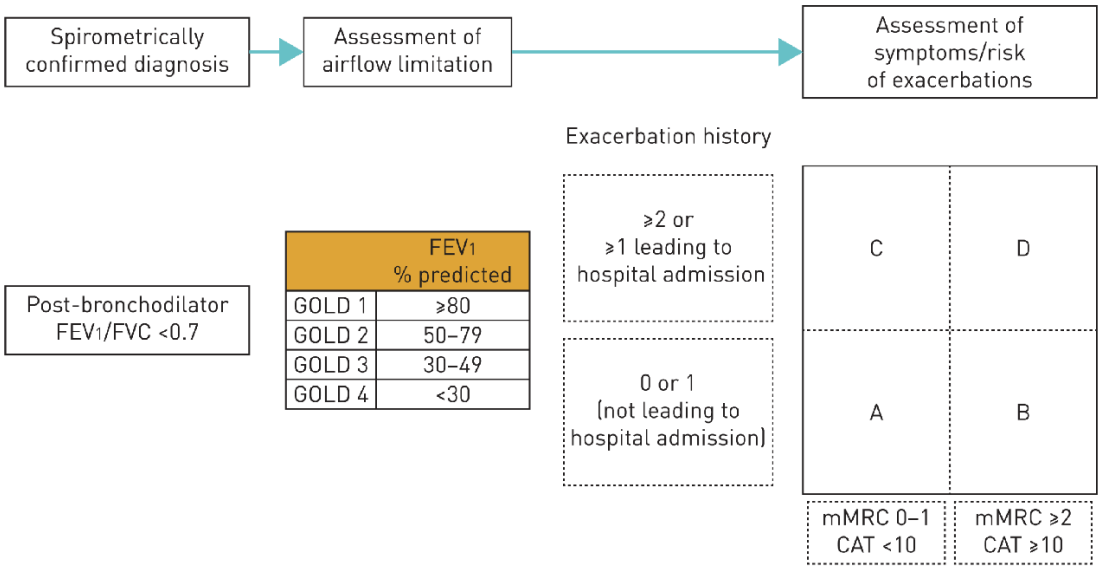

Figure 2 | Diagnosis and management of COPD according to appropriate disease severity based on degree of airflow limitation, symptom burden, and history of exacerbation. Adapted from Vogelmeier et al.[10] Reprinted with permission of the American Thoracic Society. Copyright © 2021 American Thoracic Society. All rights reserved. The American Journal of Respiratory and Critical Care Medicine is an official journal of the American Thoracic Society.

GOLD has advanced understanding of COPD over the years since its very first recommendations in 2001 from a simple assessment using degree of airflow limitation only to a more complex one

including among others exacerbation risk and symptom burden.[6,10] The evolution of the recommendations was a necessity as COPD with emerging evidence has proven to be a highly complex heterogenous disease.[18,19] While risk stratification and management of COPD has undergone significant modifications, we have not witnessed a parallel evolution of the diagnostic approach. Although current clinical practice has proven its applicability for many years, it still has some shortcomings. An important one is that COPD should only be considered in a patient that already has symptoms, thereby ignoring the possibility that mild disease may be asymptomatic.[20] A likely consequence will be delayed diagnosis and hence a missed opportunity for early prevention and treatment, which may lead to a more severe untreatable disease. Another one is whether airflow limitation should be an obligate diagnostic feature, as individuals with normal spirometry have shown evidence of airway and alveolar abnormalities usually associated with COPD.[21-26] Since COPD develops gradually over many years, these individuals could be suspected to have early disease not yet evidenced by airflow limitation.[4,7]

Undiagnosed COPD

Despite clear recommendations for when and how COPD should be suspected and diagnosed appropriately, it is unfortunately characterised by substantial underdiagnosis (Figure 3).[27] Individuals with undiagnosed COPD have been found to constitute a considerable burden to the healthcare system.[28-32] In the Canadian Cohort Obstructive Lung Disease, individuals with undiagnosed COPD still experience exacerbations despite of having fewer symptoms and less impaired lung function than those with diagnosed COPD.[30] In the National Health and Nutrition Examination Survey, although individuals with undiagnosed COPD seem healthy, they display an increased risk of early death.[31] Reported determinants of COPD underdiagnosis are younger age,

less reporting of symptoms, and milder airflow limitation.[27] Thus, individuals with undiagnosed COPD often seem to show signs of mild or perhaps early disease.

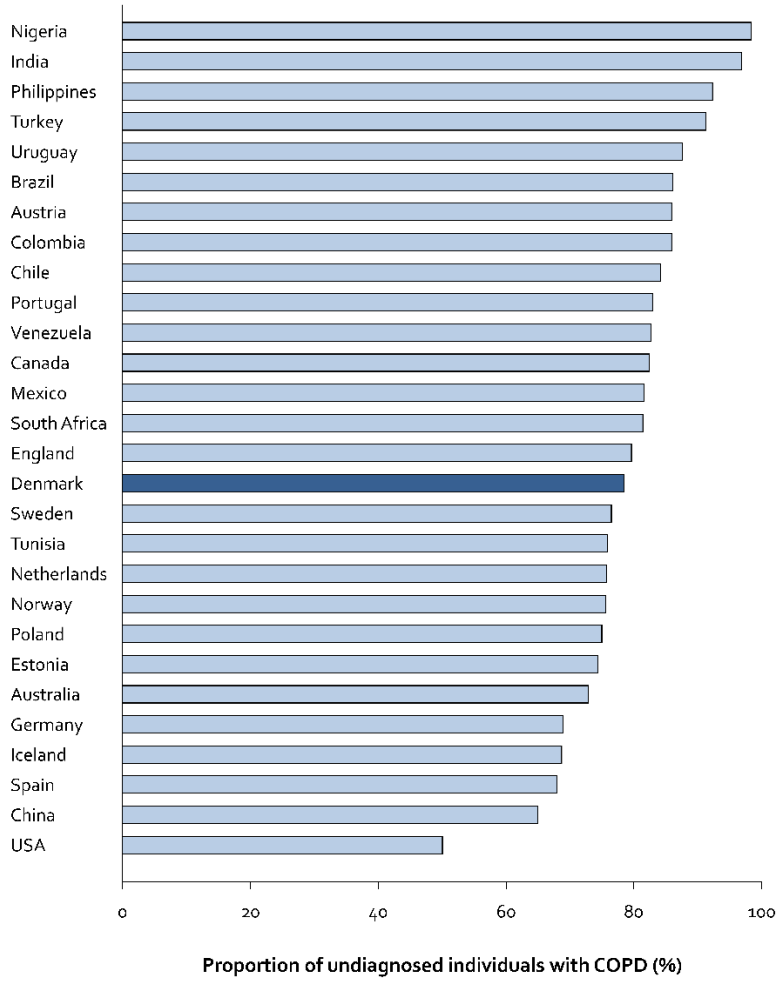

Figure 3 | Proportion of undiagnosed individuals with COPD in different countries. Numbers adapted from Lamprecht et al.[27] Number for Denmark adapted from Çolak et al.[33]

The US Preventive Services Task Force (USPSTF) recommend against screening for COPD in asymptomatic smokers, mainly because such screening did not seem to increase smoking cessation rates, but instead encourages clinicians to pursue active case-finding strategies in symptomatic smokers.[34-36] In addition, GOLD only recommends use of spirometry for early detection in symptomatic smokers.[10] However, it is difficult not to imagine that presence of clinically significant

disease may be asymptomatic as seen in other chronic diseases.[20] Also, the impact of symptoms on the prognosis of individuals with undiagnosed COPD was previously unknown. We therefore investigated the prognosis of asymptomatic and symptomatic individuals with undiagnosed COPD in the general population, and hypothesised that individuals with undiagnosed COPD would have poor prognosis compared to individuals without COPD, irrespective of presence of respiratory symptoms.[33]

Among 32 518 high-risk smokers aged ≥40 years with ≥10 pack-years of tobacco consumption in the Copenhagen General Population Study, 11% were identified with COPD, of whom 78% were undiagnosed (Figure 3). Among undiagnosed individuals with COPD, 71% were symptomatic and 29% were asymptomatic. During up to 11 years of follow-up (median: 6.1 years), 800 exacerbations, 2038 pneumonias, and 2789 deaths were registered. Compared to individuals without COPD but with similar amount of smoking exposure, individuals with undiagnosed, asymptomatic COPD displayed an increased risk of hospitalisations due to exacerbation of COPD and pneumonia, whereas individuals with undiagnosed, symptomatic COPD additionally displayed an increased risk of all-cause mortality (Figures 4 and 5). Hazard ratios (HRs) in individuals with undiagnosed, asymptomatic COPD were 4.57 (95% confidence interval [CI]: 2.59-8.06) for exacerbation, 1.64 (95% CI: 1.26-2.13) for pneumonia, and 1.20 (95% CI: 0.99-1.47) for all-cause mortality (Figure 5). Corresponding HRs in those with undiagnosed, symptomatic COPD were 12.6 (95% CI: 8.82-17.9), 2.52 (95% CI: 2.16-2.94), and 1.75 (95% CI: 1.56-1.95), respectively. Taken together, both symptomatic and asymptomatic individuals with undiagnosed COPD seem to have a poor prognosis, thereby highlighting the need for better initiatives for early diagnosis and treatment.

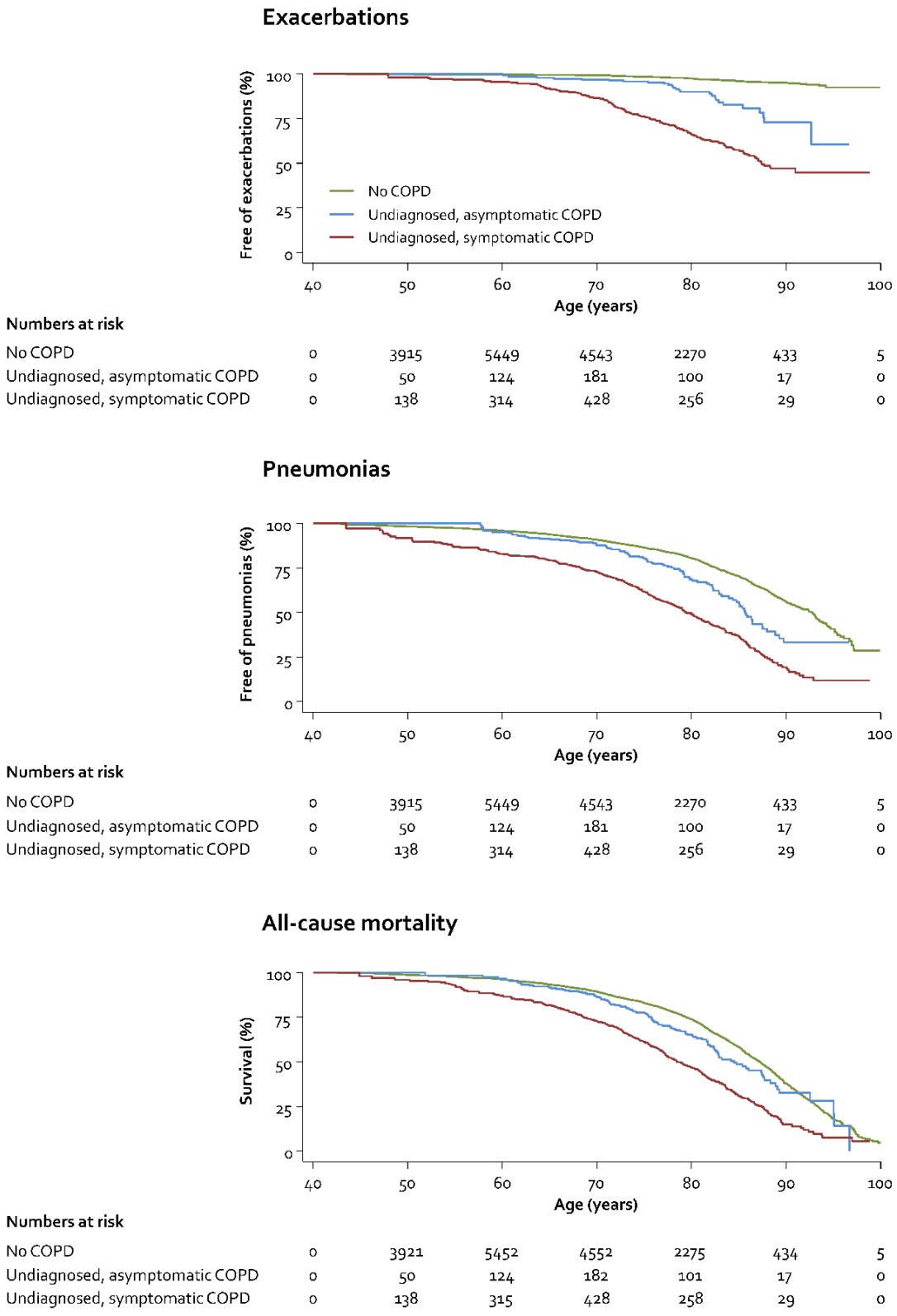

Figure 4 | Exacerbations, pneumonias, and all-cause mortality in individuals with asymptomatic and symptomatic undiagnosed COPD. Adapted from Çolak et al.[33]

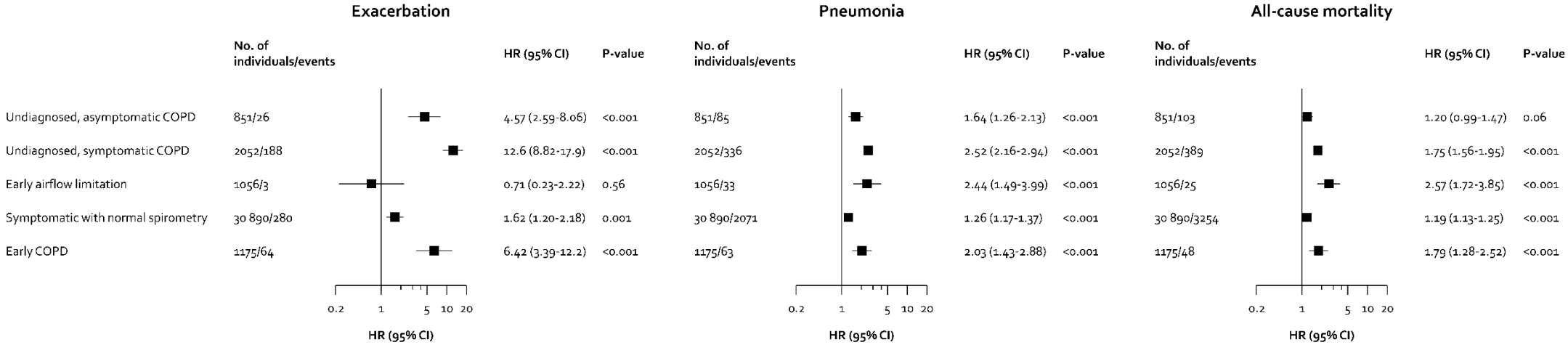

Figure 5 | Risk of exacerbation, pneumonia, and all-cause mortality in individuals with unrecognised lung disease. HRs with 95% CIs are multivariable adjusted and for undiagnosed COPD adapted from Çolak et al.[33], for early airflow limitation adapted from Çolak et al.[37] (depicted as potentially underdiagnosed airflow limitation), for symptomatic with normal spirometry adapted from Çolak et al.[38], and for early COPD adapted from Çolak et al.[39]

Underutilization of spirometry may be one of the main reasons for underdiagnosis.[27] Measurement of lung function in primary care is not prioritised as highly as measurement of blood pressure or blood cholesterol.[40,41] Number of visits to the general practitioner in the past 12 months for individuals in the Copenhagen General Population Study did not differ substantially between those without COPD and those with undiagnosed COPD,[33] suggesting that opportunities to diagnose COPD early are being missed. However, it is also well-known that patients with COPD often underreport the true burden of symptoms and may adapt their activity level according to some symptoms such as dyspnoea.[42] Underdiagnosis is likely reflecting a combination of underutilization of spirometry in general practice and patients not seeking their general practitioners despite presence of symptoms.[43] Nonetheless, a large proportion of patients with COPD in general practice do not receive treatment at first diagnosis despite having clinically significant disease,[44] suggesting that other unknown factors may affect clinicians understanding of COPD and decision to begin treatment.

A large subgroup that comprises individuals with milder airflow limitation, i.e. $FEV_1 \geq 80\%$ of predicted, was excluded as we intended to focus on clinically significant disease in order to obtain a high specificity for COPD and hence to identify most cases correctly. Another important subgroup that was excluded comprises never-smokers with COPD, which may account for one-fourth of all COPD cases in Western societies including Denmark.[45-54] Never-smokers with COPD report less symptoms and have milder airflow limitation compared to smokers with COPD, but they still display an increased risk of exacerbation and pneumonia in the Copenhagen General Population Study.[53,55] Underdiagnosis may be higher in never-smokers, as COPD is not usually considered in individuals without any smoking history. Since mild COPD and never-smokers with COPD were not considered in our study, the total burden of undiagnosed COPD is therefore expected to be underestimated.

Approximately 29% of undiagnosed individuals with COPD reported to be asymptomatic and did not fulfil the criteria for a clinical diagnosis of COPD according to GOLD or for active case-finding as suggested by the USPSTF.[10,35] Individuals with undiagnosed, asymptomatic COPD despite milder airflow limitation compared to those with undiagnosed, symptomatic COPD or diagnosed COPD still displayed an increased risk of exacerbations and pneumonias compared to those without COPD, suggesting that symptoms may not be an obligate feature to suspect and diagnose clinical COPD. Without diagnosis or intervention, it is likely that mild COPD with high treatment potential in these undiagnosed individuals will progress to a much more severe COPD with lower treatment potential. Indeed, smoking intervention will have less effect in severe compared to in mild COPD.[56] The burden of undiagnosed COPD is substantial and likely to continue with current clinical practice.

Early airflow limitation

Airflow limitation defined as a fixed ratio of post-bronchodilator $FEV_1/FVC < 0.70$ is recommended in current clinical practice to confirm the diagnosis of COPD in individuals with respiratory symptoms and relevant exposure according to GOLD.[10] Nonetheless, over the years, the appropriate definition of airflow limitation has been heatedly debated. On the one hand, the use of a fixed ratio is a simple approach to use in clinical practice, while on the other hand, it does not account for the normal age-related decline in lung function, which leads to a lower FEV_1/FVC with increasing age. A presumed consequence has therefore been potential overdiagnosis in older and underdiagnosis in younger individuals.[3,57-59] The latter situation has especially been worrisome as mild or early COPD may be overlooked, thereby delaying early prevention and treatment.

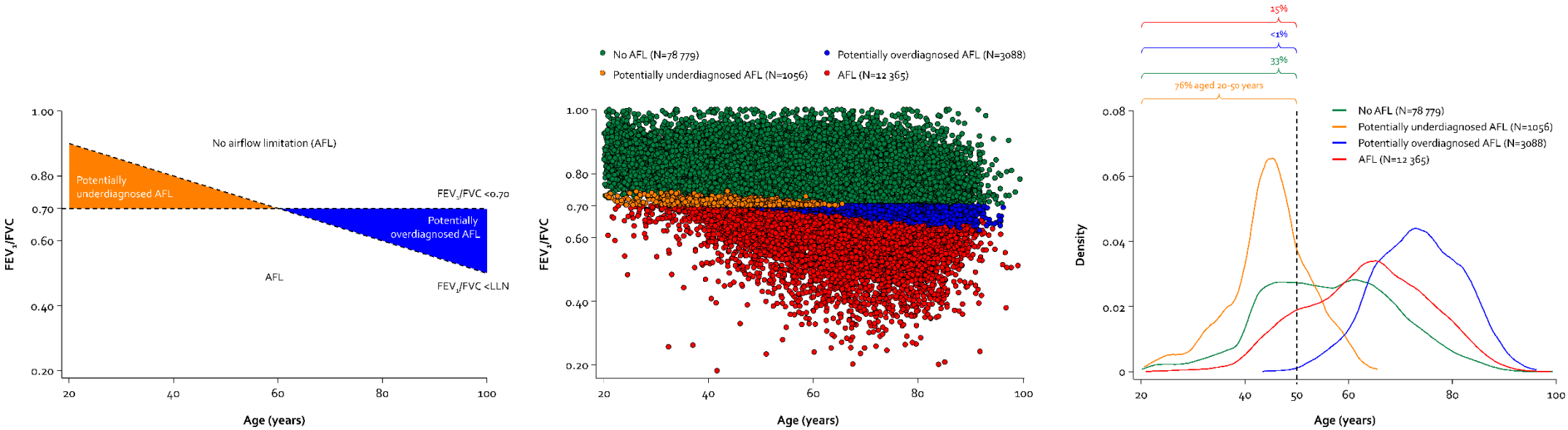

Figure 6 | Definition of airflow limitation and differences between the fixed ratio and lower limit of normal (LLN) criteria. Adapted from Çolak et al.[37] *Left panel*: A theoretical depiction of the clinical groups formed by using the fixed ratio and LLN criteria for airflow limitation. *Middle panel*: Assignment of individuals from the Copenhagen General Population Study to the formed clinical groups. *Right panel*: Age distribution of the formed clinical groups.

Instead, it has been recommended to use the lower 5th percentile of the predicted value for FEV_1/FVC, i.e. FEV_1/FVC <lower limit of normal (LLN), as such an approach will not be prone to age-related lung function decline and will also account for other biological differences (Figure 6, left panel).[60,61] However, airflow limitation as FEV_1/FVC <LLN is often defined using different lung function reference equations due to lack of standardisation. Although major efforts were made towards standardisation with the introduction of the Global Lung Initiative (GLI) lung function reference equations,[62] which is now considered as the standard given the comprehensiveness of the sampling and analyses underlying the derivation of the equations, no study has previously investigated differences between various LLN criteria relative to GOLD criteria to diagnose airflow limitation against clinical outcomes of COPD. We therefore investigated head-to-head whether GOLD (FEV_1/FVC <0.70) and four LLN criteria (FEV_1/FVC <LLN) to diagnose airflow limitation differ in identifying high-risk individuals with COPD in the general population.[63] LLN criteria used lung function reference equations from GLI, National Health and Nutrition Examination Survey (NHANES), European Community for Steel and Coal (ECSC), and Copenhagen City Heart Study (CCHS)/Copenhagen General Population Study (CGPS), respectively.

Among 108 246 individuals from the Copenhagen General Population Study, 18 111 (17%) had airflow limitation according to GOLD, 9308 (9%) according to GLI, 11 221 (10%) according to NHANES, 8855 (8%) according to ECSC, and 15 529 (14%) had airflow limitation according to CCHS/CGPS. All four LLN criteria identified more individuals with airflow limitation at younger age and fewer individuals at older age compared to GOLD, but substantial differences also existed between the different LLN criteria (Figures 7, upper panel). However, individuals identified with airflow limitation according to all five criteria reported more often chronic respiratory symptoms and higher smoking exposure compared to those without airflow limitation.

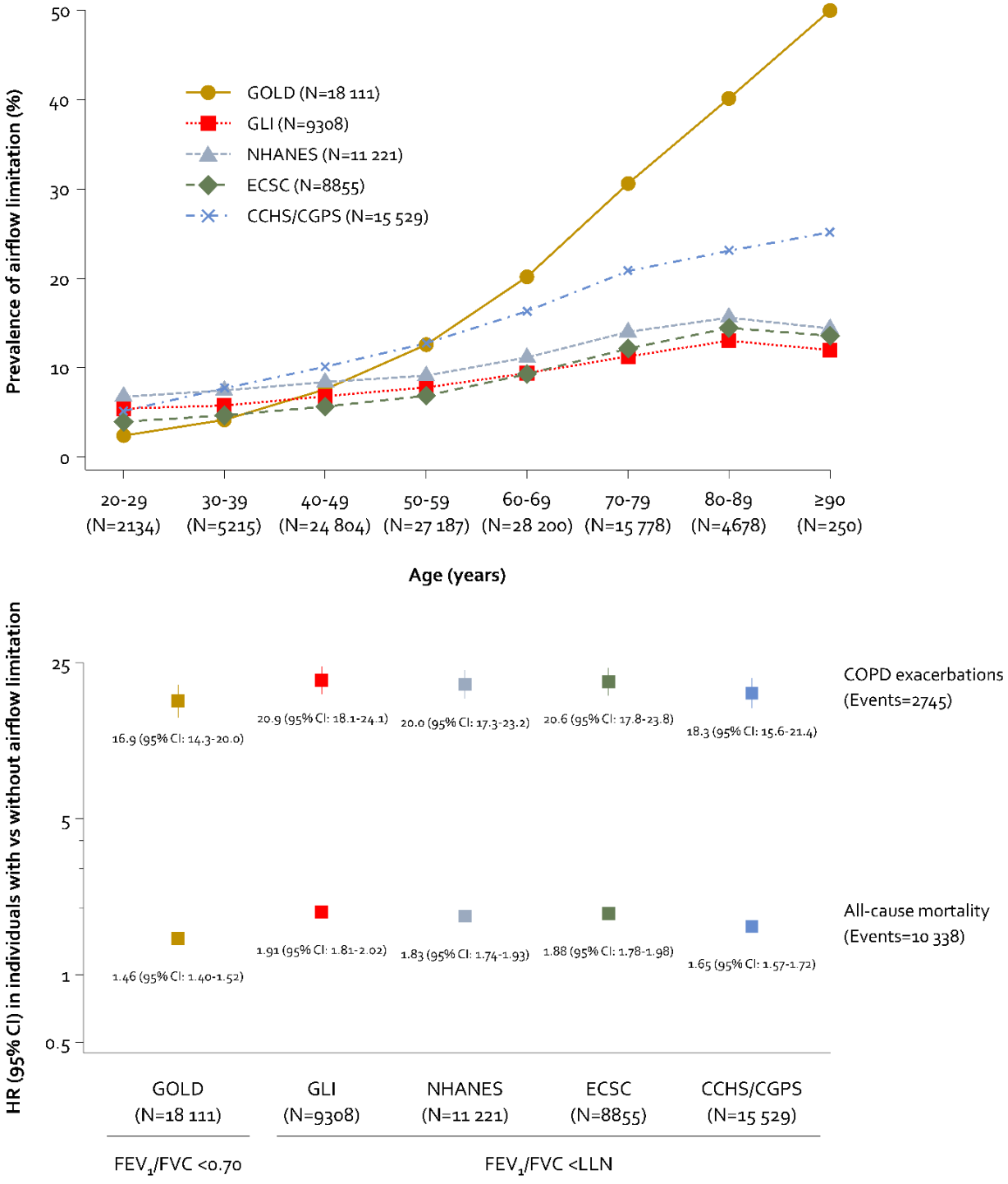

Figure 7 | Prevalence of airflow limitation and risk of COPD exacerbations and all-cause mortality in individuals with versus without airflow limitation according to five different criteria. Adapted from Çolak et al.[63]

During up to 14 years of follow-up (median: 8.7 years), 2745 COPD exacerbations and 10 338 deaths were registered. Individuals with airflow limitation according to GOLD or one of the four LLN criteria had increased risk of COPD exacerbations and all-cause mortality compared to those without airflow limitation (Figure 7, lower panel). Compared to individuals without airflow limitation, HR for COPD exacerbations in individuals with airflow limitation was 16.9 (95% CI: 14.3-20.0) according to GOLD, 20.9 (95% CI: 18.1-24.1) according to GLI, 20.0 (95% CI: 17.3-23.2) according to NHANES, 20.6 (95% CI: 17.8-23.8) according to ECSC, and 18.3 (95% CI: 15.6-21.4) according to CCHS/CGPS. Corresponding HRs for all-cause mortality were 1.46 (95% CI: 1.40-1.52), 1.91 (95% CI: 1.81-2.02), 1.83 (95% CI: 1.74-1.93), 1.88 (95% CI: 1.78-1.98), and 1.65 (95% CI: 1.57-1.72), respectively. Predictive capability for risk of COPD exacerbation based on Harrell's C statistic was slightly higher for CCHS/CGPS compared to GOLD and the other three LLN criteria, while GOLD did not differ compared to the other three LLN criteria.[63] In contrast, predictive capability for risk of all-cause mortality was slightly lower for GOLD compared to the four LLN criteria. However, all increments in Harrell's C statistics were small and did not display clinically important difference. Taken together, while the prevalence of airflow limitation ranges from 8% to 17% using GOLD and four different LLN criteria, identified individuals with airflow limitation according to all five criteria had similar risk of COPD exacerbation and all-cause mortality.

There was a large overlap of individuals when GOLD was compared to each of the four LLN criteria, but also some discordant clinical groups.[63] Among individuals diagnosed with airflow limitation according to GOLD (corresponding to N=18 111), 50% also fulfilled the diagnosis of airflow limitation according to GLI (N=9105); however, GLI only identified an additional 1% with airflow limitation (N=203). When GOLD was compared to the other three LLN criteria, corresponding proportions were 56% (N=10 565) and 4% (N=656) for NHANES, 48% (N=8745)

and 1% (N=110) for ECSC, and 74% (N=14 361) and 6% (N=1168) for CCHS/CGPS. A direct comparison of GOLD with each of the four LLN criteria showed that individuals with airflow limitation according to GOLD but not LLN criteria had increased risk of COPD exacerbations and all-cause mortality. Among the four LLN criteria, only NHANES and CCHS/CGPS identified additional individuals at risk of COPD exacerbations and all-cause mortality independent from GOLD. Compared to individuals without airflow limitation according to GOLD or the four LLN criteria, individuals with airflow limitation according to GOLD but not each of the four LLN criteria seemed older and reported more often chronic respiratory symptoms and higher smoking exposure. In contrast, individuals with airflow limitation according each of the four LLN but not GOLD criteria seemed younger, predominantly females, and reported more often chronic respiratory symptoms and asthma. Interestingly, while those identified with airflow limitation according to GOLD but not each of the four LLN criteria had 600-800 mL lower FEV_1, those identified with airflow limitation according each of the four LLN but not GOLD criteria had either 20 mL higher or 60 mL lower FEV_1 (lowest for those with airflow limitation according to NHANES and CCHS/CGPS). Thus, low FEV_1 may be the main driver for the increased risk of COPD exacerbations and all-cause mortality in discordant groups.

Individuals diagnosed with airflow limitation according to fixed ratio but not LLN have signs of clinical COPD and display an increased risk of COPD exacerbations and all-cause mortality.[58,59,63-71] Choosing LLN instead of fixed ratio for the diagnosis of airflow limitation among older individuals may lead to potential underdiagnosis of an important clinical group that will likely benefit from COPD treatment. In addition, most of the evidence for management of COPD including randomised controlled trials has only been based on airflow limitation according to fixed ratio.[10] A fixed ratio therefore seems reasonable for diagnosing airflow limitation in COPD. Nonetheless, younger individuals diagnosed with airflow limitation according to LLN but not fixed

ratio may be another important clinical group subjected to potential underdiagnosis of early airflow limitation.[63] Thus, we investigated the prognosis of potentially underdiagnosed airflow limitation in younger individuals from the general population, as defined by LLN but not fixed ratio, and hypothesised that potential underdiagnosis of airflow limitation at younger age is associated with poor prognosis.[37]

Among 95 288 individuals from the Copenhagen General Population Study, 78 779 (83%) individuals did not have airflow limitation ($FEV_1/FVC \geq 0.70$ and $\geq$LLN), 1056 (1%) had potentially underdiagnosed airflow limitation ($FEV_1/FVC \geq 0.70$ and <LLN), 3088 (3%) had potentially overdiagnosed airflow limitation ($FEV_1/FVC < 0.70$ and $\geq$LLN), and 12 365 (13%) individuals had definite airflow limitation ($FEV_1/FVC < 0.70$ and <LLN) (Figure 6, left and middle panels). In individuals with potentially underdiagnosed airflow limitation, 76% were aged 20-50 years (Figure 6, right panel). The corresponding proportion was 33% in individuals without airflow limitation. Compared to individuals without airflow limitation, individuals with potentially underdiagnosed airflow limitation were younger (median age 56 versus 45 years) and more often active smokers (28% versus 15%) but difference in tobacco consumption was small (median 14 versus 16 pack-years). After taking age into account and compared to individuals without airflow limitation, individuals with potentially underdiagnosed airflow limitation reported more often asthma (5% versus 10%), chronic respiratory symptoms (40% versus 54%), and airway medication use (4% versus 9%). Furthermore, 80% of individuals with potentially underdiagnosed airflow limitation had mild airflow limitation with $FEV_1 \geq 80\%$ predicted.

During up to 11 years of follow-up (median: 6.0 years), 2073 exacerbations, 4487 pneumonias, 3859 ischaemic heart disease events, 2046 heart failures, and 5260 deaths were registered. Individuals with potentially underdiagnosed airflow limitation had an increased risk of pneumonias

and all-cause mortality but not of exacerbations of obstructive lung disease compared to individuals without airflow limitation. Compared to individuals without airflow limitation, HR for individuals with potentially underdiagnosed airflow limitation was 0.71 (95% CI: 0.23-2.22) for exacerbation, 2.44 (95% CI: 1.49-3.99) for pneumonia, and 2.57 (95% CI: 1.72-3.85) for all-cause mortality (depicted as early airflow limitation in Figure 5). Interestingly, these individuals also displayed an increased risk of heart failure with a HR of 2.11 (95% CI: 1.09-4.10) but not of ischaemic heart disease with a HR of 0.92 (95% CI: 0.59-1.46). Taken together, younger adults identified with early airflow limitation according to LLN but not fixed ratio seem to display an increased risk of respiratory and cardiovascular morbidity and poor survival.

Individuals with airflow limitation according to the LLN but not fixed ratio seem to be a clinical group with impaired respiratory health.[37,65,69,72,73] In the European Community Respiratory Health Survey, these individuals had an increased risk of developing FEV_1 <80% predicted and hospital service utilisation due to breathing problems after 9 years of follow-up.[72] In the third National Health and Nutrition Examination Survey, the same clinical group displayed nominally the highest risk for all-cause mortality after 18 years of follow-up, despite of only comprising 20 individuals and a P-value ≥0.05.[69] In the Copenhagen General Population Study, these individuals displayed an increased risk of pneumonia, heart failure, and all-cause mortality.[37] It may be necessary to combine LLN with fixed ratio among younger individuals to identify an important clinical group with early airflow limitation at risk of COPD that would otherwise be overlooked by using fixed ratio alone. A likely explanation for presence of early airflow limitation may be underdeveloped lungs (sometimes denoted as small lungs), most likely due to low maximal attained lung function in early adulthood.[16] Low maximally attained lung function in early adulthood has been demonstrated to be an important risk factor for development of COPD later in life despite of normal age-related lung function decline.[15] Level of maximal attained lung function in early adulthood may depend on multiple

factors that influence normal lung development including genetics, prenatal exposures, birth weight, exposure to parental smoking, frequency of respiratory tract infections, and presence of asthma.[17] Interestingly, only 12% of individuals with early airflow limitation reported to have asthma in the Copenhagen General Population Study,[37] suggesting that other risk factors may be in play. Furthermore, these individuals also displayed an increased risk of heart failure. Heart failure has long been associated with airflow limitation and it can often be a diagnostic challenge in patients with COPD.[74-76] By using LLN among young individuals, we may not only be able to identify those at risk of developing COPD later in life but perhaps also those developing heart failure later in life.

Normal spirometry with chronic respiratory symptoms

A prodromal phase of COPD was once suggested by GOLD to identify high-risk smokers, previously designated as GOLD stage 0.[6] GOLD stage 0 included individuals with chronic cough and phlegm but with a normal spirometry, defined as $FEV_1/FVC \geq 0.70$. Due to incomplete evidence of whether or not individuals with GOLD stage 0 progress to GOLD stage 1 or beyond, i.e. $FEV_1/FVC < 0.70$, GOLD recommended that GOLD stage 0 should no longer be included in the diagnosis and management of COPD.[7,8] Nonetheless, smokers with normal spirometry but with chronic respiratory symptoms still constitute a significant proportion of the clinical consultations and may be at risk of developing COPD later in life.[21-24,26] We therefore investigated the prognostic significance of chronic respiratory symptoms in individuals with normal spirometry without known airway disease and tested the hypothesis that chronic respiratory symptoms are associated with respiratory hospitalisations and death in individuals with normal spirometry without known airway disease.[38] Normal spirometry was defined as $FEV_1/FVC \geq 0.70$ and chronic respiratory symptoms included dyspnoea, chronic mucus hypersecretion, wheezing, and/or cough.

Among 108 246 individuals from the Copenhagen General Population Study, 10 291 (10%) were excluded due to known airway disease, i.e. COPD and/or asthma. Among the remaining 97 955 individuals, 52 999 (54%) had normal spirometry without chronic respiratory symptoms, 30 890 (32%) had normal spirometry with chronic respiratory symptoms, 7076 (7%) had airflow limitation without chronic respiratory symptoms, and 6990 (7%) had airflow limitation with chronic respiratory symptoms. Individuals with normal spirometry with versus without chronic respiratory symptoms were older (median age 58 versus 55 years), had slightly lower lung function (median FEV_1 96% versus 101% predicted), were more active smokers (22% versus 11%) with a higher tobacco consumption (median 18 versus 11 pack-years), and had more often non-pulmonary diseases, including cardiovascular disease (13% versus 7%), diabetes (6% versus 3%), and cancer (7% versus 6%). Furthermore, these individuals reported greater healthcare use with frequent episodes of acute bronchitis/pneumonia and visits to the physician's office. Dyspnoea and wheezing were the most frequent symptoms. In individuals with normal spirometry, only 1% had FEV_1/FVC <LLN in both those with and without chronic respiratory symptoms, thereby suggesting less influence of early airflow limitation.

During up to 14 years of follow-up (median: 8.8 years), 1037 exacerbations, 5743 pneumonias, and 8750 deaths were registered. In individuals with normal spirometry, those with chronic respiratory symptoms compared to those without had increased risk of hospitalisations due to exacerbation of obstructive lung disease and pneumonia and increased all-cause mortality after adjustment for potential confounders of pulmonary and non-pulmonary diseases. HR for individuals with normal spirometry with versus without chronic respiratory symptoms was 1.62 (95% CI: 1.20-2.18) for exacerbation, 1.26 (95% CI: 1.17-1.37) for pneumonia, and 1.19 (95% CI: 1.13-1.25) for all-cause mortality (Figure 5). Increased risks could already be observed after only 2 years of follow-up, and the risk estimates were stable throughout the whole 14 years follow-up period.[38]

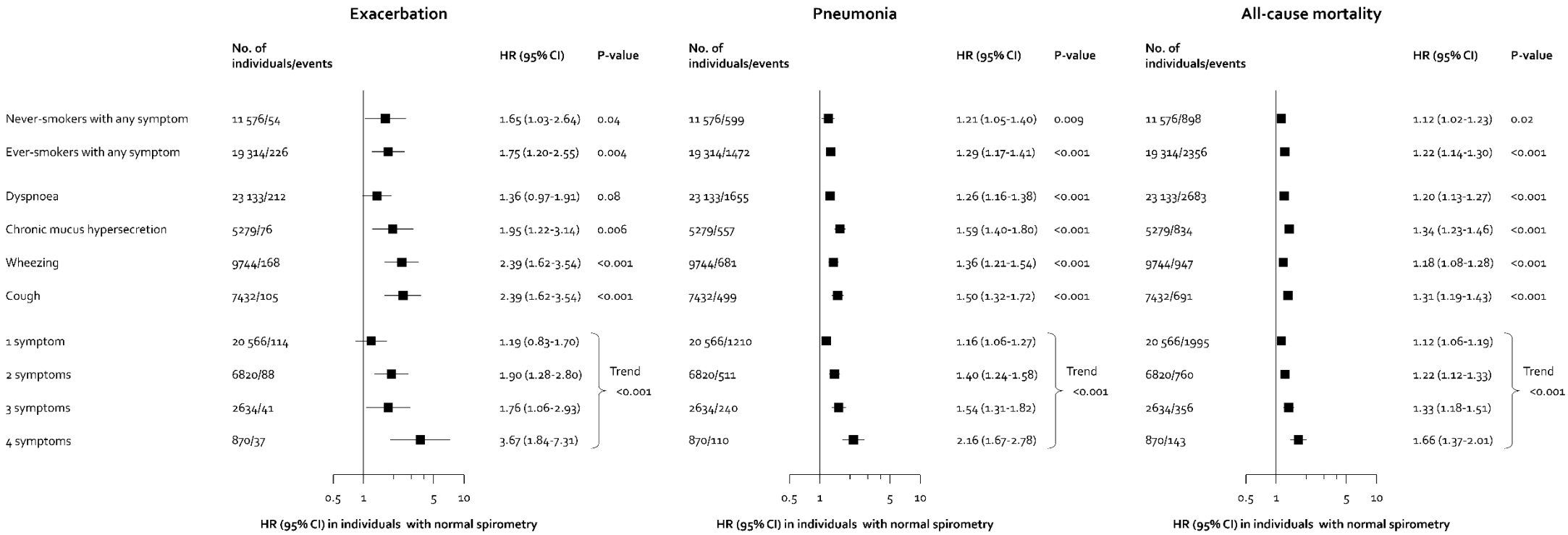

Figure 8 | Risk of exacerbation, pneumonia, and all-cause mortality in individuals with normal spirometry. HRs with 95% CIs are multivariable adjusted adapted from Çolak et al.[38]

Interestingly, results were similar when never- and ever-smokers were investigated separately (Figure 8). All types of chronic respiratory symptoms were associated with all outcomes with comparable risk estimates in individuals with normal spirometry; however, the 95% CIs for dyspnoea overlapped with 1.0 for risk of exacerbation (Figure 8). A clear dose-response relationship was present in the form of higher risk estimates with higher number of symptoms (Figure 8).

Among individuals with normal spirometry without known airway disease, as high as 32% report chronic respiratory symptoms. Even after adjustment for relevant pulmonary and non-pulmonary disease related risk factors, these individuals still display an increased risk of exacerbations of obstructive lung disease, pneumonias, and early death, including deaths with respiratory disease as underlying cause.[38] Similarly, in the Subpopulations and Intermediate Outcome Measures in COPD Study, smokers reporting severe symptoms compared to smokers reporting mild symptoms had an increased risk of COPD related exacerbations despite presence of normal spirometry.[26] In addition, symptoms, primarily in the form of chronic cough and phlegm, have been associated with increased risk of early death in individuals with normal spirometry before.[77-81] While symptoms have been associated with accelerated lung function decline and development of airflow limitation,[8,82-90] increased risks in the Copenhagen General Population Study could already be observed after only 2 years of follow-up. It is very unlikely that these individuals during this short follow-up period developed airflow limitation, especially when only 1% of them had early airflow limitation defined as FEV_1/FVC <LLN. However, in other studies, individuals with normal spirometry and chronic respiratory symptoms have shown evidence of airway and alveolar abnormalities, including increased airway wall thickness, pulmonary emphysema, gas trapping, and abnormal diffusing capacity, well-known attributes of clinical COPD.[23,25,26,91-94] It is therefore reasonable to suspect early COPD in these individuals not yet evidenced by airflow limitation.[4] However, it seems that

only a small proportion of them progress over time from having symptoms alone to comorbid airflow limitation.[8] Thus, symptoms in the absence of airflow limitation may not only be a marker of early COPD but perhaps also impaired respiratory health.[7] Another important consideration is whether current clinical practice would ever lead to a diagnosis of COPD in these individuals, as they may never develop airflow limitation but still display signs of COPD. At the moment, this has also become the main reason for discussing whether airflow limitation should remain an obligate diagnostic feature for COPD.[95,96]

Early COPD

An international group of experts has recently proposed an operational definition for early COPD by distinguishing "early" from "mild" disease.[4] While mild COPD indicates that the disease has already developed and focus should be on treatment to halt progression, early COPD means that the disease has not fully developed and may still be prevented. Accordingly, early COPD should be defined in individuals aged <50 years with a smoking exposure ≥10 pack-years with one or more of the following: (i) early airflow limitation defined as post-bronchodilator FEV_1/FVC <LLN, (ii) compatible thoracic computed tomography (CT) abnormalities such as visual emphysema, air trapping, and/or bronchial thickening graded mild or worse, and/or (iii) evidence of accelerated FEV_1 decline relative to FVC such as ≥60 mL/year. Hitherto, no information has been available on the impact of early COPD and knowledge has, as an alternative, been extrapolated from what is known as mild COPD.[97] We therefore investigated the prevalence, characteristics, and prognosis of individuals with early COPD in the general population.[39]

Among 105 630 individuals from the Copenhagen General Population Study, 8064 (8%) were aged <50 years with a smoking exposure ≥10 pack-years, of whom 1175 (15%) had early COPD, defined

as FEV_1/FVC <LLN. Among these 8064 high-risk young smokers, early COPD was prevalent in 7% of those aged 20-29 years, 10% of those aged 30-39 years, and 15% of those aged 40-49 years (Figure 9). More than two-third of individuals with early COPD reported at least one chronic respiratory symptom. Prevalence of FEV_1/FVC <LLN was lower among those with smoking exposure <10 pack-years.

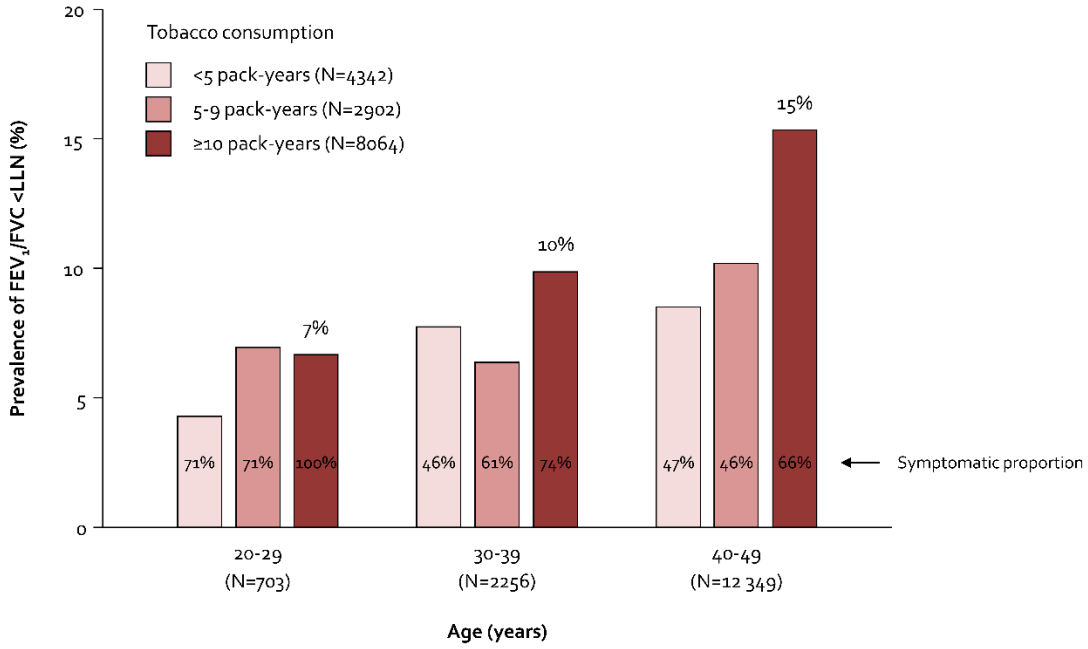

Figure 9 | Prevalence of FEV_1/FVC <LLN according to age and smoking exposure. Adapted from Çolak et al.[39] Reprinted with permission of the American Thoracic Society. Copyright © 2021 American Thoracic Society. All rights reserved. The American Journal of Respiratory and Critical Care Medicine is an official journal of the American Thoracic Society.

Compared to high-risk young smokers at risk but without early COPD, those with early COPD were more often active smokers (46% versus 58%) and had a higher tobacco consumption (median 19 versus 23 pack-years). A substantial proportion also had significant lung function impairment with FEV_1/FVC <0.70 (1% versus 75%) and FEV_1 <80% of predicted (9% versus 40%), probably due to the LLN criteria being included in the operational definition for early COPD.

During up to 14 years of follow-up (median: 10 years), we observed 117 exacerbations of obstructive lung disease, 227 pneumonias, and 185 deaths among these 8064 high-risk young smokers. Compared to individuals without early COPD, HR in individuals with early COPD was 6.42 (95% CI: 3.39-12.2) for exacerbation, 2.03 (95% CI: 1.43-2.88) for pneumonia, and 1.79 (95% CI: 1.28-2.52) for all-cause mortality (Figure 5). When individuals with and without early COPD were stratified according to presence of chronic respiratory symptoms, symptomatic individuals without early COPD had an increased risk of exacerbations and pneumonias but not of all-cause mortality, whereas symptomatic individuals with early COPD had an increased risk of all investigated outcomes. Asymptomatic individuals with and without early COPD did not differ regarding prognosis. No evidence of interaction between early COPD and chronic respiratory symptoms was found. Individuals with early COPD more often display chronic respiratory symptoms and severe lung function impairment, and an increased risk of obstructive lung disease and pneumonia related hospitalisations and early death.

Upon comparison, younger individuals with early COPD displayed lower relative risk estimates for exacerbations but higher for pneumonias and all-cause mortality compared to older individuals with COPD (in this instance defined as FEV_1/FVC <LLN in those aged ≥50 years with smoking exposure ≥10 pack-years).[39] Interestingly, differences were smaller when comparing the symptomatic subgroups of younger individuals with early COPD and older individuals with COPD. In contrast, asymptomatic older individuals with COPD had poorer prognosis than asymptomatic younger individuals with early COPD, probably due to more lung function impairment with FEV_1 <80% predicted (24% in younger versus 40% in older individuals).

More than half of individuals with early COPD reported to be active smokers and hence available for smoking intervention that could potentially halt progression of lung damage and change the

disease course accordingly. Importantly, a substantial proportion already had clinical signs of disease onset at the baseline examination in the form of chronic respiratory symptoms and significant lung function impairment with FEV_1/FVC <0.70 and FEV_1 <80% of predicted. An increased risk of exacerbations and pneumonias was already observed after a very short follow-up time. The newly proposed operational definition for early COPD therefore seems to capture not only mild but also moderate cases of COPD. A refinement of the definition may therefore seem warranted so the very early phases of disease development can be captured in order to implement prevention before irreversible lung damage has taken place.

Approximately half of individuals without early COPD reported chronic respiratory symptoms. Symptomatic individuals with and without early COPD had at the baseline examination many clinical features in common and shared similar poor prognosis. It is possible that symptomatic individuals despite displaying $FEV_1/FVC \geq LLN$ may still have other abnormalities as evidenced by thoracic CT and/or FEV_1 decline suggesting presence of early COPD. The high burden of early COPD in our study is therefore likely underestimated, as the other requirements to fulfil the operational definition for early COPD were not applied.

Asymptomatic individuals with early COPD shared similar baseline characteristics and did not differ in prognosis compared to asymptomatic individuals without early COPD, suggesting that presence of chronic respiratory symptoms may be a marker of disease progression when applying the operational definition for early COPD.[4] Despite lung function impairment, asymptomatic individuals with early COPD may perhaps not develop clinical COPD and could easily represent individuals with impaired respiratory health only, probably due to low maximal attained lung function in early adulthood.[98]

From early to clinical COPD

The burden of early COPD seems substantial, but whether young adults identified with early COPD will develop clinical COPD later in life is unknown.[39] In addition, many individuals identified with early COPD already show signs of clinical COPD that will likely benefit from treatment. Since never-smokers and smokers with low smoking exposure such as <10 pack-years may also develop COPD,[46,51,53,99,100] progression from early to clinical COPD would also be relevant to investigate in these subgroups even if they do not fulfil the threshold for smoking exposure in the operational definition for early COPD.[4] Future patients with COPD will likely have less smoking exposure due to the decreasing smoking prevalence and increasing proportion of smokers with low tobacco consumption.[101] We therefore investigated risk of clinical COPD 10 years later in young adults from the general population with and without early COPD with a focus on smoking exposure.[55]

Among 89 054 adults with $FEV_1/FVC \geq 0.70$ at baseline examination from the Copenhagen General Population Study, 14 870 had lung function measurement at the final examination 10 years later. Among these 14 870 individuals, 5497 were aged <50 years at the baseline examination, of whom 168 (3%) had early COPD, defined as FEV_1/FVC <LLN (corresponding to early airflow limitation as $FEV_1/FVC \geq 0.70$ and <LLN). At the final examination 10 years later, out of 5497 individuals, 104 (2%) had developed clinical COPD, defined as chronic respiratory symptoms with FEV_1/FVC <0.70 and FEV_1 <80% predicted.

During follow-up, individuals with versus without early COPD did not differ regarding FEV_1 decline (median 23 versus 22 mL/year). While FEV_1 decline ≥60 mL/year was observed in 8% of those with early COPD, it was observed in 6% of those without early COPD. In contrast, individuals that developed clinical COPD during follow-up had an FEV_1 decline of 45 mL/year

(31% with FEV_1 decline ≥60 mL/year), whereas individuals without clinical COPD had an FEV_1 decline of 21 mL/year (6% with FEV_1 decline ≥60 mL/year).

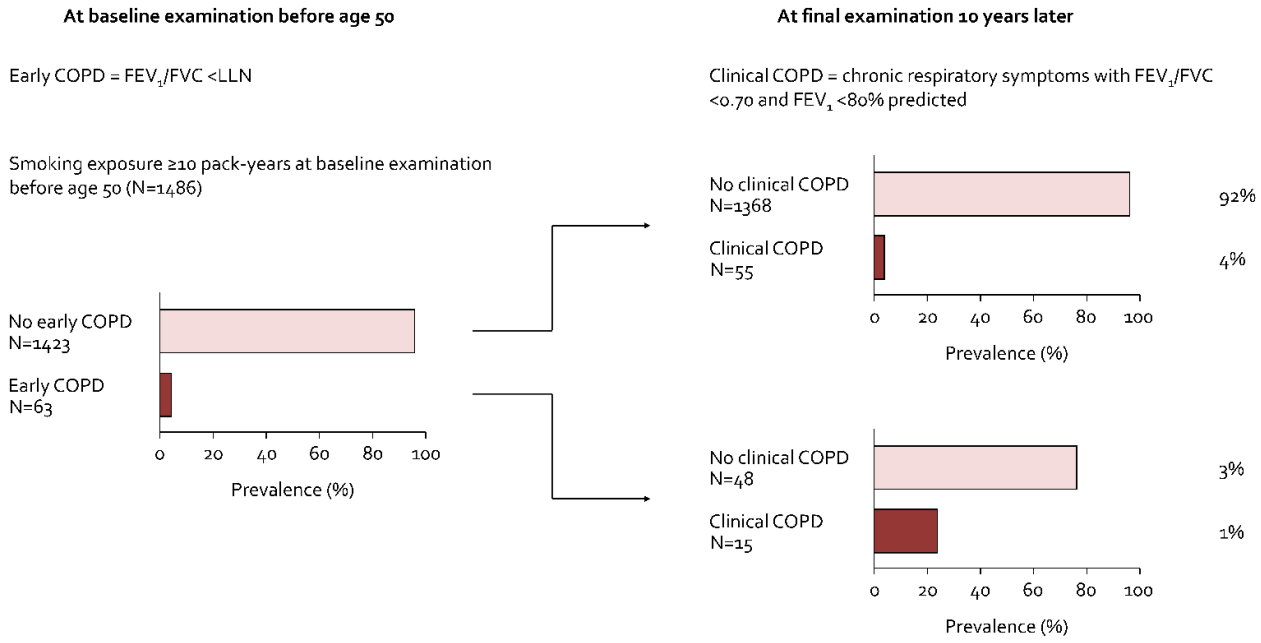

Figure 10 | Early COPD at baseline examination before age 50 and development of clinical COPD at final examination 10 years later. Adapted from Çolak et al.[55] Reprinted with permission of the American Thoracic Society. Copyright © 2021 American Thoracic Society. All rights reserved. The American Journal of Respiratory and Critical Care Medicine is an official journal of the American Thoracic Society.

At the baseline examination, prevalence of early COPD in individuals aged <50 years was 4% in smokers with ≥10 pack-years, 3% in smokers with <10 pack-years, and 2% in never-smokers. Thus, after exclusion of individuals with FEV_1/FVC <0.70 at the baseline examination, 4% and not 15% seemed to fulfil the operational definition for early COPD in the Copenhagen General Population Study.[4,39,55] At the final examination 10 years later in smokers with ≥10 pack-years, 24% developed clinical COPD in those with early COPD versus 4% in those without early COPD (Figure 10). Corresponding numbers were 10% and 1% in smokers with <10 pack-years, and 3% and <1% in never-smokers, respectively.

Choosing a smoking exposure threshold as ≥10 pack-years at the baseline examination in individuals with early COPD yielded a sensitivity of 24%, a specificity of 96%, a positive predictive value of 21%, and a negative predictive value of 97% for predicting clinical COPD at the final examination 10 years later. Changing smoking exposure from ≥10 pack-years to ignoring the quantity, i.e. including all smokers, dropped sensitivity slightly from 24% to 18% without any noteworthy change in specificity, or in positive- or negative predictive values. Sensitivity dropped to 13% when never-smokers were also included without any large change in the other values.

Risk of clinical COPD at the final examination 10 years later in individuals with early COPD at the baseline examination did not differ substantially with odds ratios (ORs) of 7.77 (95% CI: 4.10-14.7) in smokers with ≥10 pack-years and 8.56 (95% CI: 4.92-14.9) in all smokers.[55] Risk of exacerbation with obstructive lung disease during follow-up also did not differ with corresponding HRs of 4.16 (95% CI: 1.66-10.5) and 4.33 (95% CI: 1.89-9.93), respectively. Results were independently validated in the Copenhagen City Heart Study.

Depending on amount of smoking exposure, less than 24% of individuals defined with early COPD at the baseline examination developed clinical COPD at the final examination 10 years later, thereby demonstrating how difficult it is to define early COPD and the potential for overdiagnosis in younger adults. Nonetheless, a high negative predictive value of 97% was obtained, suggesting that the proposed operational definition for early COPD may be good at excluding individuals not likely to develop clinical COPD later in life. In fact, less than 4% of individuals with normal spirometry subsequently developed clinical COPD.

Smoking exposure ≥10 pack-years has been suggested as a requirement in the recently proposed operational definition for early COPD.[4] However, younger individuals seem to be less represented in those with smoking exposure ≥10 pack-years when age distribution was investigated in

individuals aged <50 years at the baseline examination.[55] Among all 5497 individuals in the Copenhagen General Population Study, 20% were aged 20-39 and 80% were aged 40-49. Corresponding numbers were 25% and 75% in those with FEV_1/FVC <LLN, 14% and 86% in those with smoking exposure ≥10 pack-years, and 16% and 84% in those with both FEV_1/FVC <LLN and smoking exposure ≥10 pack-years, respectively. Since it will usually take a longer time for an average smoker to obtain a higher tobacco consumption, the age distribution moves towards higher age. Capturing the very early phases of disease development in COPD will therefore be reduced by having a high smoking exposure threshold in the definition. Never-smokers and smokers with low tobacco consumption may also develop COPD and will not benefit from early prevention and intervention with the present operational definition for early COPD.

Clinical perspectives and future investigations

Emerging evidence now suggests the existence of individuals with early COPD, but not all of them seem to progress over time to develop clinical COPD.[39,55] Therefore, if all with early COPD are identified and treated, an unfortunate consequence will be potential overdiagnosis of a large group subjected to preventive measures to halt progression and development of COPD. In consequence, identification of markers for disease progression from early to clinical COPD is a necessity in future investigations to avoid overdiagnosis. So far, combination of early airflow limitation with chronic respiratory symptoms seems promising. Nonetheless, individuals identified with early COPD still display impaired respiratory health by experiencing all well-known complications of COPD including early death.

It is important to note that we do not have the necessary evidence to support that intervention in individuals with early COPD halts progression and development of COPD later in life. Effect of a

certain intervention can therefore only be speculated on by extrapolating knowledge from available trials performed in individuals with established COPD. Since many individuals with early COPD report to be active smokers,[39,55] an important intervention would nonetheless be smoking cessation. Smoking cessation should be regarded as the most effective preventive measure for COPD.[10] Indeed, in the Lung Health Study, a randomised controlled trial involving individuals with mild to moderate COPD, smoking cessation was not only associated with lower FEV_1 decline and fewer respiratory symptoms but also improved overall survival.[56,102-104] By intervening in the very early phases of disease development in COPD before irreversible lung damage has taken place, a greater effect of smoking cessation may be anticipated. Furthermore, since individuals with early COPD displayed increased risk of exacerbations of obstructive lung disease and early death like individuals with developed COPD, it would only be reasonable to contemplate on pharmacological therapies targeting COPD, including LABA, LAMA, and ICS. A meta-analysis of randomised controlled trials testing the effect of pharmacological therapies in individuals with developed COPD demonstrated that active therapy versus placebo attenuated FEV_1 decline with 5 mL/year (7 mL/years when restricting to ICS).[105] These effect sizes were likely underestimates as approximately half of the included subjects were in the sixth decade of life, where the necessary FEV_1 decline to develop COPD has already taken place. Indeed, in the UPLIFT study, individuals with COPD aged <50 years yielded an FEV_1 decline difference of 20 mL/year between active therapy versus placebo.[106] Since treatment response seems higher when therapy is initiated at younger age, and presuming that therapy can be initiated in individuals with early COPD, the putative minimal clinically important difference of 100 mL can already be achieved after only 5 years of treatment.[107] Other interventions that would be relevant to test in individuals with early COPD include pneumococcal and influenza vaccination due to the increased risk of pneumonias.

COPD has its origin in early life and develops gradually over many years. Prejudice in the form of COPD being a self-inflicted disease by smoking in older individuals should be abandoned. Development of COPD will not be eliminated with absence of or low smoking exposure, which will become increasingly evident in the future with worldwide decreasing smoking prevalence.[101] Current clinical practice is challenged with severe untreatable COPD due to late diagnosis. Addressing individuals at high risk of developing COPD may be part of the solution, which unfortunately is not prioritised in current clinical practice. Like the cardiovascular risk factors hypertension and hypercholesterolemia, lung function impairment and chronic respiratory symptoms should perhaps also be regarded as risk factors needing attention. Lung function impairment and chronic respiratory symptoms comprise pathological conditions associated with increased morbidity and mortality, mostly with COPD as underlying cause. An obvious question for future research is therefore whether available treatments for COPD could be relevant to target also in individuals with these risk factors. Some may argue that the consequences of overtreatment may be less worrisome than the consequences of developing COPD; however, this also needs to be tested in randomised controlled trials. Overtreatment of individuals with hypertension and hypercholesterolemia is already taking place to prevent cardiovascular disease, that is, a large fraction of individuals receiving antihypertensive and cholesterol-lowering treatment do not personally benefit with reduced cardiovascular disease, but a crucial difference is that the needed evidence is provided to understand the balance between over- versus undertreatment. Since available treatments for COPD have been proven to be effective, should the indication be further expanded to include individuals with early COPD at risk? As mentioned above, only large randomised controlled trials involving individuals at high risk of developing COPD and focusing on primary prevention can provide the necessary evidence.

Concluding remarks

Current clinical practice is challenged by the increased burden of severe untreatable COPD, highlighting the need for early diagnosis before irreversible lung damage has taken place. Since COPD has its origin early in life and develops gradually over many years, individuals that will develop COPD could be identified before disease onset allowing for implementation of preventive measures thereby likely halting progression. Although airflow limitation in the clinical diagnosis of COPD is defined as $FEV_1/FVC < 0.70$, a subgroup of younger individuals predisposed to develop COPD later in life can be identified with early airflow limitation defined as $FEV_1/FVC < LLN$. Younger individuals identified with early airflow limitation display an increased risk of hospitalisations due to obstructive lung disease and an increased risk of early death. Individuals with chronic respiratory symptoms despite normal spirometry also seem to be predisposed to develop COPD. Depending on the amount of smoking exposure, prevalence of early COPD ranges up to 4%, of whom one in four develops clinical COPD 10 years later, suggesting challenges with potential overdiagnosis and overtreatment if all with early COPD were treated. On the other hand, individuals identified with early COPD have signs of impaired respiratory health with increased morbidity and mortality and could potentially benefit from COPD targeted preventive measures, which needs to be documented before implemented. In any case, an age cut-off of 50 years and smoking exposure of 10 pack-years may be too late for early diagnosis and intervention for COPD. Therefore, large randomised controlled trials investigating primary prevention in individuals at risk of developing COPD later in life are needed.

References

1. Mortality GBD, Causes of Death C. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015; **385**: 117-71.

2. Martinez FD. Early-Life Origins of Chronic Obstructive Pulmonary Disease. *N Engl J Med* 2016; **375**: 871-8.

3. Rennard SI, Drummond MB. Early chronic obstructive pulmonary disease: definition, assessment, and prevention. *Lancet* 2015; **385**: 1778-88.

4. Martinez FJ, Han MK, Allinson JP, et al. At the Root: Defining and Halting Progression of Early Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* 2018; **197**: 1540-51.

5. Celli BR, Decramer M, Wedzicha JA, et al. An official American Thoracic Society/European Respiratory Society statement: research questions in COPD. *Eur Respir J* 2015; **45**: 879-905.

6. Pauwels RA, Buist AS, Calverley PM, Jenkins CR, Hurd SS, Committee GS. 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**: 1256-76.

7. Rodriguez-Roisin R, Han MK, Vestbo J, Wedzicha JA, Woodruff PG, Martinez FJ. Chronic Respiratory Symptoms with Normal Spirometry. A Reliable Clinical Entity? *Am J Respir Crit Care Med* 2017; **195**: 17-22.

8. Vestbo J, Lange P. Can GOLD Stage 0 provide information of prognostic value in chronic obstructive pulmonary disease? *Am J Respir Crit Care Med* 2002; **166**: 329-32.

9. Agusti A, Hogg JC. Update on the Pathogenesis of Chronic Obstructive Pulmonary Disease. *N Engl J Med* 2019; **381**: 1248-56.

10. Vogelmeier CF, Criner GJ, Martinez FJ, 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**: 557-82.

11. Fletcher C, Peto R. The natural history of chronic airflow obstruction. *Br Med J* 1977; **1**: 1645-8.

12. Fletcher CM, Peto R, Tinker CM, Speizer FE. The natural history of chronic bronchitis and emphysema: an eight-year study of early chronic obstructive lung disease in working men in London: Oxford University Press; 1976.

13. Burrows B, Knudson RJ, Lebowitz MD. The relationship of childhood respiratory illness to adult obstructive airway disease. *Am Rev Respir Dis* 1977; **115**: 751-60.

14. Vestbo J, Lange P. Fletcher and Peto 40 Years On. A Tribute and Reflection. *Am J Respir Crit Care Med* 2017; **195**: 1420-2.

15. Lange P, Celli B, Agusti A, et al. Lung-Function Trajectories Leading to Chronic Obstructive Pulmonary Disease. *N Engl J Med* 2015; **373**: 111-22.

16. Agusti A, Faner R. Lung function trajectories in health and disease. *Lancet Respir Med* 2019; **7**: 358-64.

17. Postma DS, Bush A, van den Berge M. Risk factors and early origins of chronic obstructive pulmonary disease. *Lancet* 2015; **385**: 899-909.

18. Hurst JR, Vestbo J, Anzueto A, et al. Susceptibility to exacerbation in chronic obstructive pulmonary disease. *N Engl J Med* 2010; **363**: 1128-38.

19. Vestbo J, Edwards LD, Scanlon PD, et al. Changes in forced expiratory volume in 1 second over time in COPD. *N Engl J Med* 2011; **365**: 1184-92.

20. Soriano JB, Zielinski J, Price D. Screening for and early detection of chronic obstructive pulmonary disease. *Lancet* 2009; **374**: 721-32.

21. Dransfield MT, Kunisaki KM, Strand MJ, et al. Acute Exacerbations and Lung Function Loss in Smokers with and without Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* 2017; **195**: 324-30.

22. Martinez CH, Murray S, Barr RG, et al. Respiratory Symptoms Items from the COPD Assessment Test Identify Ever-Smokers with Preserved Lung Function at Higher Risk for Poor Respiratory Outcomes. An Analysis of the Subpopulations and Intermediate Outcome Measures in COPD Study Cohort. *Ann Am Thorac Soc* 2017; **14**: 636-42.

23. Regan EA, Lynch DA, Curran-Everett D, et al. Clinical and Radiologic Disease in Smokers With Normal Spirometry. *JAMA Intern Med* 2015; **175**: 1539-49.

24. Tan WC, Bourbeau J, Hernandez P, et al. Exacerbation-like respiratory symptoms in individuals without chronic obstructive pulmonary disease: results from a population-based study. *Thorax* 2014; **69**: 709-17.

25. Tan WC, Hague CJ, Leipsic J, et al. Findings on Thoracic Computed Tomography Scans and Respiratory Outcomes in Persons with and without Chronic Obstructive Pulmonary Disease: A Population-Based Cohort Study. *PLoS One* 2016; **11**: e0166745.

26. Woodruff PG, Barr RG, Bleecker E, et al. Clinical Significance of Symptoms in Smokers with Preserved Pulmonary Function. *N Engl J Med* 2016; **374**: 1811-21.

27. Lamprecht B, Soriano JB, Studnicka M, et al. Determinants of underdiagnosis of COPD in national and international surveys. *Chest* 2015; **148**: 971-85.

28. Bednarek M, Maciejewski J, Wozniak M, Kuca P, Zielinski J. Prevalence, severity and underdiagnosis of COPD in the primary care setting. *Thorax* 2008; **63**: 402-7.

29. Coultas DB, Mapel D, Gagnon R, Lydick E. The health impact of undiagnosed airflow obstruction in a national sample of United States adults. *Am J Respir Crit Care Med* 2001; **164**: 372-7.

30. Labonte LE, Tan WC, Li PZ, et al. Undiagnosed Chronic Obstructive Pulmonary Disease Contributes to the Burden of Health Care Use. Data from the CanCOLD Study. *Am J Respir Crit Care Med* 2016; **194**: 285-98.

31. Martinez CH, Mannino DM, Jaimes FA, et al. Undiagnosed Obstructive Lung Disease in the United States. Associated Factors and Long-term Mortality. *Ann Am Thorac Soc* 2015; **12**: 1788-95.

32. Miravitlles M, Soriano JB, Garcia-Rio F, et al. Prevalence of COPD in Spain: impact of undiagnosed COPD on quality of life and daily life activities. *Thorax* 2009; **64**: 863-8.

33. Çolak Y, Afzal S, Nordestgaard BG, Vestbo J, Lange P. Prognosis of asymptomatic and symptomatic, undiagnosed COPD in the general population in Denmark: a prospective cohort study. *Lancet Respir Med* 2017; **5**: 426-34.

34. Au DH. Screening for Chronic Obstructive Pulmonary Disease: D Is the New F. *JAMA Intern Med* 2016; **176**: 601-2.

35. Guirguis-Blake JM, Senger CA, Webber EM, Mularski RA, Whitlock EP. Screening for Chronic Obstructive Pulmonary Disease: Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA* 2016; **315**: 1378-93.

36. Martinez FJ, O'Connor GT. Screening, Case-Finding, and Outcomes for Adults With Unrecognized COPD. *JAMA* 2016; **315**: 1343-4.

37. Çolak Y, Afzal S, Nordestgaard BG, Vestbo J, Lange P. Young and middle-aged adults with airflow limitation according to lower limit of normal but not fixed ratio have high morbidity and poor survival: a population-based prospective cohort study. *Eur Respir J* 2018; **51**: 1702681.

38. Çolak Y, Nordestgaard BG, Vestbo J, Lange P, Afzal S. Prognostic significance of chronic respiratory symptoms in individuals with normal spirometry. *Eur Respir J* 2019; **54**: 1900734.

39. Çolak Y, Afzal S, Nordestgaard BG, Vestbo J, Lange P. Prevalence, Characteristics, and Prognosis of Early Chronic Obstructive Pulmonary Disease. The Copenhagen General Population Study. *Am J Respir Crit Care Med* 2020; **201**: 671-80.

40. Barr RG, Celli BR, Mannino DM, et al. Comorbidities, patient knowledge, and disease management in a national sample of patients with COPD. *Am J Med* 2009; **122**: 348-55.

41. Roberts NJ, Smith SF, Partridge MR. Why is spirometry underused in the diagnosis of the breathless patient: a qualitative study. *BMC Pulm Med* 2011; **11**: 37.

42. Miravitlles M, Ribera A. Understanding the impact of symptoms on the burden of COPD. *Respir Res* 2017; **18**: 67.

43. Jones RC, Price D, Ryan D, et al. Opportunities to diagnose chronic obstructive pulmonary disease in routine care in the UK: a retrospective study of a clinical cohort. *Lancet Respir Med* 2014; **2**: 267-76.

44. Gruffydd-Jones K, Brusselle G, Jones R, et al. Changes in initial COPD treatment choice over time and factors influencing prescribing decisions in UK primary care: in UK primary care: a real-world, retrospective, observational. *NPJ Prim Care Respir Med* 2016; **26**: 16002.

45. Bridevaux PO, Probst-Hensch NM, Schindler C, et al. Prevalence of airflow obstruction in smokers and never-smokers in Switzerland. *Eur Respir J* 2010; **36**: 1259-69.

46. Çolak Y, Afzal S, Nordestgaard BG, Lange P. Majority of never-smokers with airflow limitation do not have asthma: the Copenhagen General Population Study. *Thorax* 2016; **71**: 614-23.

47. Hagstad S, Ekerljung L, Lindberg A, Backman H, Ronmark E, Lundback B. COPD among non-smokers - report from the obstructive lung disease in Northern Sweden (OLIN) studies. *Respir Med* 2012; **106**: 980-8.

48. Lamprecht B, McBurnie MA, Vollmer WM, et al. COPD in never smokers: results from the population-based burden of obstructive lung disease study. *Chest* 2011; **139**: 752-63.

49. Menezes AM, Perez-Padilla R, Jardim JR, et al. Chronic obstructive pulmonary disease in five Latin American cities (the PLATINO study): a prevalence study. *Lancet* 2005; **366**: 1875-81.

50. Miravitlles M, Ferrer M, Pont A, et al. Characteristics of a population of COPD patients identified from a population-based study. Focus on previous diagnosis and never smokers. *Respir Med* 2005; **99**: 985-95.

51. Salvi SS, Barnes PJ. Chronic obstructive pulmonary disease in non-smokers. *Lancet* 2009; **374**: 733-43.

52. Tan WC, Sin DD, Bourbeau J, et al. Characteristics of COPD in never-smokers and ever-smokers in the general population: results from the CanCOLD study. *Thorax* 2015; **70**: 822-9.

53. Thomsen M, Nordestgaard BG, Vestbo J, Lange P. Characteristics and outcomes of chronic obstructive pulmonary disease in never smokers in Denmark: a prospective population study. *Lancet Respir Med* 2013; **1**: 543-50.

54. Zhou Y, Wang C, Yao W, et al. COPD in Chinese nonsmokers. *Eur Respir J* 2009; **33**: 509-18.

55. Çolak Y, Afzal S, Nordestgaard BG, Lange P, Vestbo J. Importance of Early COPD In Young Adults for Development of Clinical COPD: Findings from the Copenhagen General Population Study. *Am J Respir Crit Care Med* 2020; **doi:10.1164/rccm.202003-0532OC**.

56. Anthonisen NR, Connett JE, Kiley JP, et al. Effects of smoking intervention and the use of an inhaled anticholinergic bronchodilator on the rate of decline of FEV1. The Lung Health Study. *JAMA* 1994; **272**: 1497-505.

57. Çolak Y, Løkke A, Marott JL, Lange P, Vestbo J. Impact of diagnostic criteria on the prevalence of COPD. *Clin Respir J* 2013; **7**: 297-303.

58. Mannino DM, Sonia Buist A, Vollmer WM. Chronic obstructive pulmonary disease in the older adult: what defines abnormal lung function? *Thorax* 2007; **62**: 237-41.

59. Mohamed Hoesein FA, Zanen P, Lammers JW. Lower limit of normal or FEV1/FVC < 0.70 in diagnosing COPD: an evidence-based review. *Respir Med* 2011; **105**: 907-15.

60. Bakke PS, Rönmark E, Eagan T, et al. Recommendations for epidemiological studies on COPD. *Eur Respir J* 2011; **38**: 1261-77.

61. Pellegrino R, Viegi G, Brusasco V, et al. Interpretative strategies for lung function tests. *Eur Respir J* 2005; **26**: 948-68.

62. Quanjer PH, Stanojevic S, Cole TJ, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. *Eur Respir J* 2012; **40**: 1324-43.

63. Çolak Y, Nordestgaard BG, Vestbo J, Lange P, Afzal S. Comparison of five major airflow limitation criteria to identify high-risk individuals with COPD: a contemporary population-based cohort. *Thorax* 2020; **75**: 944-54.

64. Akkermans RP, Berrevoets MA, Smeele IJ, et al. Lung function decline in relation to diagnostic criteria for airflow obstruction in respiratory symptomatic subjects. *BMC Pulm Med* 2012; **12**: 12.

65. Akkermans RP, Biermans M, Robberts B, et al. COPD prognosis in relation to diagnostic criteria for airflow obstruction in smokers. *Eur Respir J* 2014; **43**: 54-63.

66. Bhatt SP, Balte PP, Schwartz JE, et al. Discriminative Accuracy of FEV1:FVC Thresholds for COPD-Related Hospitalization and Mortality. *JAMA* 2019; **321**: 2438-47.

67. Bhatt SP, Sieren JC, Dransfield MT, et al. Comparison of spirometric thresholds in diagnosing smoking-related airflow obstruction. *Thorax* 2014; **69**: 409-14.

68. Güder G, Brenner S, Angermann CE, et al. "GOLD or lower limit of normal definition? A comparison with expert-based diagnosis of chronic obstructive pulmonary disease in a prospective cohort-study". *Respir Res* 2012; **13**: 13.

69. Mannino DM, Diaz-Guzman E. Interpreting lung function data using 80% predicted and fixed thresholds identifies patients at increased risk of mortality. *Chest* 2012; **141**: 73-80.

70. Mohamed Hoesein FA, Zanen P, Sachs AP, Verheij TJ, Lammers JW, Broekhuizen BD. Spirometric thresholds for diagnosing COPD: 0.70 or LLN, pre- or post-dilator values? *COPD* 2012; **9**: 338-43.

71. Zaigham S, Wollmer P, Engstrom G. Lung function, forced expiratory volume in 1 s decline and COPD hospitalisations over 44 years of follow-up. *Eur Respir J* 2016; **47**: 742-50.

72. Cerveri I, Corsico AG, Accordini S, et al. Underestimation of airflow obstruction among young adults using FEV1/FVC <70% as a fixed cut-off: a longitudinal evaluation of clinical and functional outcomes. *Thorax* 2008; **63**: 1040-5.

73. Pothirat C, Chaiwong W, Phetsuk N, Liwsrisakun C. Misidentification of airflow obstruction: prevalence and clinical significance in an epidemiological study. *Int J Chron Obstruct Pulmon Dis* 2015; **10**: 535-40.

74. Brenner S, Güder G, Berliner D, et al. Airway obstruction in systolic heart failure--COPD or congestion? *Int J Cardiol* 2013; **168**: 1910-6.

75. Lam CS, Lyass A, Kraigher-Krainer E, et al. Cardiac dysfunction and noncardiac dysfunction as precursors of heart failure with reduced and preserved ejection fraction in the community. *Circulation* 2011; **124**: 24-30.

76. Wannamethee SG, Shaper AG, Papacosta O, Lennon L, Welsh P, Whincup PH. Lung function and airway obstruction: associations with circulating markers of cardiac function and incident heart failure in older men-the British Regional Heart Study. *Thorax* 2016; **71**: 526-34.

77. Ekberg-Aronsson M, Pehrsson K, Nilsson JA, Nilsson PM, Lofdahl CG. Mortality in GOLD stages of COPD and its dependence on symptoms of chronic bronchitis. *Respir Res* 2005; **6**: 98.

78. Lange P, Nyboe J, Appleyard M, Jensen G, Schnohr P. Relation of ventilatory impairment and of chronic mucus hypersecretion to mortality from obstructive lung disease and from all causes. *Thorax* 1990; **45**: 579-85.

79. Mannino DM, Buist AS, Petty TL, Enright PL, Redd SC. Lung function and mortality in the United States: data from the First National Health and Nutrition Examination Survey follow up study. *Thorax* 2003; **58**: 388-93.

80. Stavem K, Sandvik L, Erikssen J. Breathlessness, phlegm and mortality: 26 years of follow-up in healthy middle-aged Norwegian men. *J Intern Med* 2006; **260**: 332-42.

81. Stavem K, Sandvik L, Erikssen J. Can global initiative for Chronic Obstructive Lung Disease stage 0 provide prognostic information on long-term mortality in men? *Chest* 2006; **130**: 318-25.

82. Allinson JP, Hardy R, Donaldson GC, Shaheen SO, Kuh D, Wedzicha JA. The Presence of Chronic Mucus Hypersecretion across Adult Life in Relation to Chronic Obstructive Pulmonary Disease Development. *Am J Respir Crit Care Med* 2016; **193**: 662-72.

83. Brito-Mutunayagam R, Appleton SL, Wilson DH, Ruffin RE, Adams RJ, North West Adelaide Cohort Health Study T. Global Initiative for Chronic Obstructive Lung Disease stage 0 is associated with excess FEV(1) decline in a representative population sample. *Chest* 2010; **138**: 605-13.

84. de Marco R, Accordini S, Cerveri I, et al. Incidence of chronic obstructive pulmonary disease in a cohort of young adults according to the presence of chronic cough and phlegm. *Am J Respir Crit Care Med* 2007; **175**: 32-9.

85. Guerra S, Sherrill DL, Venker C, Ceccato CM, Halonen M, Martinez FD. Chronic bronchitis before age 50 years predicts incident airflow limitation and mortality risk. *Thorax* 2009; **64**: 894-900.

86. Kalhan R, Dransfield MT, Colangelo LA, et al. Respiratory Symptoms in Young Adults and Future Lung Disease. The CARDIA Lung Study. *Am J Respir Crit Care Med* 2018; **197**: 1616-24.

87. Lindberg A, Jonsson AC, Ronmark E, Lundgren R, Larsson LG, Lundback B. Ten-year cumulative incidence of COPD and risk factors for incident disease in a symptomatic cohort. *Chest* 2005; **127**: 1544-52.

88. Mejza F, Gnatiuc L, Buist AS, et al. Prevalence and burden of chronic bronchitis symptoms: results from the BOLD study. *Eur Respir J* 2017; **50**.

89. Perez-Padilla R, Fernandez-Plata R, Montes de Oca M, et al. Lung function decline in subjects with and without COPD in a population-based cohort in Latin-America. *PLoS One* 2017; **12**: e0177032.

90. Vestbo J, Prescott E, Lange P. Association of chronic mucus hypersecretion with FEV1 decline and chronic obstructive pulmonary disease morbidity. Copenhagen City Heart Study Group. *Am J Respir Crit Care Med* 1996; **153**: 1530-5.

91. Bodduluri S, Reinhardt JM, Hoffman EA, et al. Signs of Gas Trapping in Normal Lung Density Regions in Smokers. *Am J Respir Crit Care Med* 2017; **196**: 1404-10.

92. Coxson HO, Leipsic J, Parraga G, Sin DD. Using pulmonary imaging to move chronic obstructive pulmonary disease beyond FEV1. *Am J Respir Crit Care Med* 2014; **190**: 135-44.

93. Harvey BG, Strulovici-Barel Y, Kaner RJ, et al. Risk of COPD with obstruction in active smokers with normal spirometry and reduced diffusion capacity. *Eur Respir J* 2015; **46**: 1589-97.

94. Tsushima K, Sone S, Yoshikawa S, et al. Clinical differences in the Global Initiative for Chronic Obstructive Lung Disease Stage 0. *Respir Med* 2006; **100**: 1360-7.

95. Barnes PJ, Vestbo J, Calverley PM. The Pressing Need to Redefine "COPD". *Chronic Obstr Pulm Dis* 2019; **6**: 380-3.

96. Lowe KE, Regan EA, Anzueto A, et al. COPDGene((R)) 2019: Redefining the Diagnosis of Chronic Obstructive Pulmonary Disease. *Chronic Obstr Pulm Dis* 2019; **6**: 384-99.

97. Soriano JB, Polverino F, Cosio BG. What is early COPD and why is it important? *Eur Respir J* 2018; **52**: doi: 10.1183/13993003.01448-2018.

98. Agusti A, Noell G, Brugada J, Faner R. Lung function in early adulthood and health in later life: a transgenerational cohort analysis. *Lancet Respir Med* 2017; **5**: 935-45.

99. Çolak Y, Lange P. The long-lasting dark shadow of past and present smoking. *Lancet Respir Med* 2020; **8**: 3-5.

100. Oelsner EC, Balte PP, Bhatt SP, et al. Lung function decline in former smokers and low-intensity current smokers: a secondary data analysis of the NHLBI Pooled Cohorts Study. *Lancet Respir Med* 2020; **8**: 34-44.

101. Ng M, Freeman MK, Fleming TD, et al. Smoking prevalence and cigarette consumption in 187 countries, 1980-2012. *JAMA* 2014; **311**: 183-92.

102. Anthonisen NR, Skeans MA, Wise RA, et al. The effects of a smoking cessation intervention on 14.5-year mortality: a randomized clinical trial. *Ann Intern Med* 2005; **142**: 233-9.

103. Kanner RE, Connett JE, Williams DE, Buist AS. Effects of randomized assignment to a smoking cessation intervention and changes in smoking habits on respiratory symptoms in smokers

with early chronic obstructive pulmonary disease: the Lung Health Study. *Am J Med* 1999; **106**: 410-6.

104. Scanlon PD, Connett JE, Waller LA, et al. Smoking cessation and lung function in mild-to-moderate chronic obstructive pulmonary disease. The Lung Health Study. *Am J Respir Crit Care Med* 2000; **161**: 381-90.

105. Celli BR, Anderson JA, Cowans NJ, et al. Pharmacotherapy and Lung Function Decline in Patients with Chronic Obstructive Pulmonary Disease: A Systematic Review. *Am J Respir Crit Care Med* 2020; **doi:10.1164/rccm.202005-1854OC**.

106. Morice AH, Celli B, Kesten S, Lystig T, Tashkin D, Decramer M. COPD in young patients: a pre-specified analysis of the four-year trial of tiotropium (UPLIFT). *Respir Med* 2010; **104**: 1659-67.

107. Jones PW, Beeh KM, Chapman KR, Decramer M, Mahler DA, Wedzicha JA. Minimal clinically important differences in pharmacological trials. *Am J Respir Crit Care Med* 2014; **189**: 250-5.

Part II – Methods with a critical review

Study design and populations

The Copenhagen General Population Study (CGPS) is a Danish contemporary population-based cohort initiated on November 26, 2003 with ongoing enrolment.[1-6] Since the initiation of the study, >150 000 individuals have been examined. In Denmark, all individuals are assigned a unique identification number (the Central Person Registration number) at birth or immigration and recorded with information such as date and place of birth, sex, residency, citizenship, and date of death in the national Danish Civil Registration System since its establishment in 1968.[7-9] Individuals aged 20-100 years living in the Capital Region of Denmark are randomly selected from the national Danish Civil Registration System to reflect the adult Danish general population (response-rate 43%). Approximately 25% of the eligible individuals aged <40 years are randomly selected, whereas all eligible individuals aged ≥40 years are randomly selected. Participants are invited by a letter, and if they do not respond, a second letter is sent. All participants complete a questionnaire, undergo a physical examination, and give blood for biochemical and genetic analyses. Questionnaires are reviewed at the day of attendance by a healthcare professional together with the participant. The study was approved by Herlev and Gentofte Hospital and a Danish ethical committee (identification number: H-KF-01-144/01) and was conducted according to the Declaration of Helsinki. All participants provide written informed consent. We have included individuals with complete information on lung function at baseline examination recruited from November 26, 2003, to April 28, 2015. An ongoing follow-up examination was initiated in March 31, 2014, where individuals are invited systematically based on region and previous participation date following similar recruitment criteria as described above, thereby allowing an approximately follow-up time of 10 years for individuals participating twice.[1,10]

The Copenhagen City Heart Study (CCHS) is a Danish historical population-based cohort initiated in 1976-78 with follow-up examinations in 1981-83, 1991-94, 2001-03, and 2011-13, recruited and examined as the CGPS but from different parts of Copenhagen.[1,11] The CCHS was used for independent external validation of some of the findings in the CGPS.[1] In this regard, no individual appeared in more than one study. Since we needed a comparable time of follow-up as in the CGPS, we used information on lung function and smoking exposure from the 1981-83, 1991-94, and 2001-03 examinations in order to approximate a follow-up time of 10 years, i.e. we followed individuals from 1981-83 through 1991-94, and individuals from 1991-94 through 2001-03.

An important type of bias to consider in a population-based cohort is selection bias, which arises due to a systematic error in the recruitment process.[12] The consequences of selection bias will be that the estimated study parameters in the cohort are not truly representative of the general population, thereby limiting external validity or generalisability. However, individuals in the CGPS were sampled from the national Danish Civil Registration System, where all individuals in Denmark are registered, and study findings can therefore easily be extrapolated to the Danish general population, especially due to the large sample-size where random variation is redundant. Differential selection according to exposure and outcome is also less likely, as individuals in the CGPS were randomly sampled before outcome onset. Nonetheless, we have only sampled a certain proportion of individuals aged <40 years, which may influence generalisability in younger individuals. It is also important to note that individuals were sampled from the Capital Region, which despite of comprising the largest proportion of individuals living in Denmark may influence generalisability to other Danish regions; however, since Denmark is a relatively small homogeneous country, this may be less important.

Non-response is a potential source of bias that should be considered, since we had a response-rate of 43% in the CGPS. Non-response bias arises due to systematic differences between non-responders and responders. Generalisability will be affected if non-responders differ from responders regarding exposure and outcome. However, this type of bias often does not affect exposure and outcome due to random sampling, i.e. internal validity, which means that the tested hypotheses between exposure and outcome are valid.[12] Nonetheless, it is well-known that non-responders compared to responders are characterised by a more severe disease phenotype.[12] Non-response bias may therefore have underestimated the true burden of disease in individuals with early undiagnosed COPD in the CGPS compared to the Danish general population, as only the healthiest individuals respond to the invitation and participate.

Losses to follow-up is another important source of bias in population-based cohorts, especially when individuals that are lost during follow-up are systematically different from those who remain in the study, which may again influence generalisability.[12] However, we did not lose track of even a single individual during follow-up in the CGPS due to the unique Central Person Registration number provided to everyone at birth or immigration via the national Danish Civil Registration System. Individuals who emigrated during follow-up were censored at the date of emigration, which in the CGPS comprised 0.4% (452 out of 108 246 with complete information on lung function). It is therefore unlikely that losses to follow-up can be considered as a bias in the CGPS, and even if present, it probably has minimal importance.

COPD

Defining chronic obstructive pulmonary disease (COPD) has always been a challenge in clinical epidemiology.[13] Although individuals identified with COPD in a population-based cohort cannot be

equated to patients diagnosed with COPD in clinical practice, we can always approximate a clinical diagnosis of COPD by defining a typical average patient with COPD. Almost all patients with COPD in clinical practice are diagnosed in accordance with the Global Initiative for Chronic Obstructive Lung Disease (GOLD) recommendations by confirming airflow limitation, defined as a fixed ratio of post-bronchodilator forced expiratory volume in 1 second (FEV_1)/forced expiratory volume (FVC) <0.70, in those with respiratory symptoms and a relevant exposure.[14] Worldwide, tobacco smoking remains the single most important risk factor for COPD,[14] with the exception that some never-smokers due to other forms of exposure also develop COPD.[15] By combining lung function with information on respiratory symptoms and smoking exposure in the CGPS, we will be approximating a clinical diagnosis of COPD. Additional sensitivity analyses can be carried out by using stricter criteria for defining COPD, which will increase specificity but reduce sensitivity for the clinical diagnosis of COPD. That said, these prerequisites for COPD diagnosis do not necessarily apply to what has become known as early COPD, as this is an unknown field, where we mostly rely on hypotheses and operational definitions based on the natural history of airflow limitation and COPD.[16,17]

Lung function

All randomly invited participants in the CGPS completed spirometry at the physical examination without any prior selection criteria. Spirometry was conducted by trained healthcare professionals according to the standard operating procedure for spirometry performance developed specifically for the CGPS, which has undergone a rigorous validation process before.[3,18] Pre-bronchodilator measurements of FEV_1 and FVC were performed, typically measured with at least three sets of values and up to seven. Spirometry was performed in a standing position without the use of a nose-

clip under strict instructions from the trained healthcare professional. A valid spirometry performance was based on at least two measurements differing by less than 5% and a correct visual inspection of the spirometry curves. Only the highest measurements of FEV_1 and FVC were used.

Spirometry can often be challenging. Measurement errors often arise due to inadequate or incomplete inhalation, a slow start or lack of blast effort during exhalation, additional breath taken during manoeuvre, lack of tight lips around the mouthpiece, early stop of exhalation, some exhalation through the nose, and/or coughing.[19] FVC will often be more affected than FEV_1. In order to minimise such errors, observation during performance and visual inspection of the spirometry curves are necessary, all of which are incorporated in the standard operating procedure for spirometry performance in the CGPS.[18] More than three measurements are usually taken if performance errors are observed or suspected. Using only the highest measurements of the spirometric indices will also likely reduce such type of errors. On rare occasions, participants are really incapable of performing an adequate spirometry usually due to old age and/or a chronic condition, where we here instead have prioritised the presence of an invalid measurement rather than the absence of a valid measurement.[20,21] Here, we simply acknowledge the fact that old age and chronic conditions will affect spirometry performance and lung function, which indeed is more realistic and reflective of clinical practice. Despite such potential measurement problems in some individuals, a large sample-size in the CGPS ensures a proper random distribution of lung function measurements including those with and without measurement error thereby reflecting the Danish general population.

In the first 14 625 participants, spirometry was performed using a Vitalograph Spirometer (Maids Moreton, Buckinghamshire, UK), and in the remaining participants, it was performed using an EasyOne Spirometer (ndd Medical Technologies, Zurich, Switzerland).[3,18] The Vitalograph was

replaced in 2005 as it stopped functioning. Therefore, we do not have data allowing comparison of measurements from the two spirometers on the same participants. However, lung function distribution was overall similar in the CGPS between Vitalograph and EasyOne measurements,[22] suggesting that large systematic differences between the two spirometers should not be considered as an issue. While Vitalograph was calibrated daily with a 1-L syringe, EasyOne was verified regularly with a 3-L syringe in accordance with the manufacturers recommendations. These two spirometers are categorised as light-weighted office spirometers and should fulfil the minimum quality criteria, but they are not as accurate as the spirometers designed for respiratory function laboratories. The spirometers were operated by multiple healthcare professionals, which may lead to a systematic difference in recording of measurement and assessment of spirometry quality. While this may be considered as a flaw in the study design, it resembles a typical clinical practice, where such types of spirometers are more frequently used and operated by multiple personnel. In order to standardise spirometry performance, the staff was trained properly using standard operating procedures in spirometry performance, which was certified on three occasions by more experienced instructors including myself YÇ.

Predicted values were calculated using internally derived reference values based on a subsample of healthy asymptomatic never-smokers without any chronic condition with age and height as covariates separately for men and women, comprising 11 288 individuals aged 20-100 years (10 572 from the CGPS and 716 from the CCHS).[3,18] Asymptomatic was defined as being without dyspnoea, chronic mucus hypersecretion, wheezing, and cough. Chronic condition was defined in accordance to the questionnaire and national Danish Patient Registry and included among others respiratory- and cardiovascular disease, diabetes, and cancer.

Airflow limitation was defined according to a fixed ratio, i.e. FEV_1/FVC <0.70, and the lower limit of normal (LLN), i.e. FEV_1/FVC <LLN; the LLN was defined as the bottom 5th percentile of the predicted value for FEV_1/FVC, calculated as the mean value minus 1.645 standard deviations.[3] Predicted values for FEV_1/FVC were usually calculated using internally derived reference values based on the CCHS and CGPS, but they were also calculated according to reference values based on (1) the Global Lung Initiative (GLI),[23] (2) the National Health and Nutrition Examination Survey (NHANES),[24] and (3) the European Community for Steel and Coal (ECSC).[25] While the fixed ratio was designated as the GOLD criteria, the other four LLN criteria were designated as CCHS/CGPS, GLI, NHANES, and ECSC criteria, respectively. It is important to note that GOLD recommends use of post-bronchodilator FEV_1/FVC <0.70 for the clinical diagnosis of COPD.[14] Since post-bronchodilator spirometry was not performed in the CGPS, a slight modification of the GOLD criteria was a necessity. However, we were also interested in whether pre-bronchodilator spirometry could identify high-risk individuals with COPD, as spirometry in general is underused in clinical practice, and if spirometry is used, the post-bronchodilator part is often omitted due to being very time-consuming. Although prevalence of airflow limitation varied as much as 8% to 17% between the five different criteria, the distribution of airflow limitation severity according to FEV_1 % of predicted was surprisingly similar including in those with FEV_1/FVC <0.70.[6] Interestingly, the fact that all the different airflow limitation criteria identified high-risk individuals with COPD to a similar degree suggests that they all have high internal validity without any clinical noteworthy differences between them. However, when we combined the different airflow limitation criteria, we witnessed only moderate degree of overlap and large size of discordant groups, where some of them had a COPD-like disease and prognosis. In this regard, it was obvious that GOLD and the locally derived LLN criteria, i.e. CCHS/CGPS, seemed optimal for identifying high-risk individuals with COPD,[6] which were mostly done in all our investigations.

We were unable to classify the type of airflow limitation as reversible or irreversible, as only pre-bronchodilator but not post-bronchodilator spirometry was performed. A reversible airflow limitation may indicate a diagnosis of asthma, and an irreversible airflow limitation may indicate a diagnosis of COPD or asthma-COPD overlap.[26] Some of the individuals identified with airflow limitation could therefore be suspected to have asthma. Although using pre-bronchodilator instead of post-bronchodilator spirometric indices have been shown to overestimate prevalence of COPD,[13,27] there seems to be no difference in diagnostic accuracy for COPD between them.[28] Since airflow limitation was defined in a high-risk population with substantial smoking exposure typically combined with moderate to severe airflow limitation with FEV_1 <80% predicted, usually applied as inclusion criteria in clinical trials with COPD and exclusion criteria in clinical trials with asthma, the majority would be expected to have COPD or asthma-COPD overlap.[26] When relevant, individuals with self-reported asthma were also excluded in sensitivity analyses and results were similar. Since we also aimed at identifying and investigating individuals with early COPD, we deliberately did not exclude individuals with asthma in all analyses, as asthma may precede and contribute to the risk of developing COPD.[16] Furthermore, individuals identified with COPD experienced frequent acute emergency department visits and/or hospitalisations with a primary diagnosis of COPD in the national Danish Patient Registry, which has previously been shown to have a high positive predictive value of 92% for the diagnosis of COPD.[29] An exacerbation of COPD could therefore be used to validate the clinical diagnosis of COPD in individuals identified with airflow limitation. Nonetheless, we may have overlooked a group of individuals with a certain degree of reversibility in lung function despite normal spirometry or absent airflow limitation that may, if reversibility is of clinical magnitude, suggest undiagnosed asthma.[30] Such type of misclassification will likely be non-differential, as reversibility will likely be present both in those

with and without airflow limitation, perhaps even more in those with airflow limitation, and hence bias towards the null-hypothesis and not likely explain the positive findings in our studies.

Respiratory symptoms

Information on respiratory symptoms in the CGPS was obtained from the questionnaire. All randomly invited participants have completed the questionnaire at home and had it reviewed at the day of attendance by a healthcare professional together with the participant. Respiratory symptoms included dyspnoea, chronic mucus hypersecretion, wheezing, and cough.[2,5] Dyspnoea was defined as breathlessness or troubled breathing during different levels of walking, at night-time, while bathing, getting dressed, seated, and/or at rest. Chronic mucus hypersecretion was defined as coughing and phlegm from the lungs in the morning and/or during the day as long as three consecutive months each year. Wheezing was defined as occasional whistling or wheezing while breathing. Cough was defined as occasionally coughing during activity.

Recall bias should not be an issue, since many of the applied respiratory symptoms did not have reference to a specific time horizon. In contrast, interviewer bias could be suspected, as all participants despite of completing the questionnaire at home will have it reviewed at the day of attendance by a healthcare professional. However, since all participants in the CGPS are randomly selected without any prior knowledge on previous medical history or condition, interviewer bias is less likely to be present. Questionnaires are also reviewed before the physical examination and blood sampling, so the healthcare professionals are completely blinded from the objective measurements.

Despite similar disease severity, individuals may differ in symptom perception.[31] While some will have high symptom perception and tend to overreport, others will have low and underreport. Whether this should be considered as a bias or simply the nature of people can be discussed. It is, however, important to be aware of these circumstances and accept that individuals will inevitably differ in how symptoms are acknowledged and reported. Since we ought to investigate or identify an average patient with symptoms, the unusual cases that will disturb a true association between exposure and outcome, e.g. asymptomatic individuals with severe airflow limitation, will only become a statistical noise that will diminish, as these individuals will be expected to comprise a very small proportion in an otherwise large random sampling. Furthermore, the consequence will likely be non-differential misclassification biasing towards the null-hypothesis for later developed morbidity or mortality.

Since some of the used questions on respiratory symptoms in the CGPS are activity-dependent, mostly those related to dyspnoea, the responses may be biased if individuals adapt their activity level to reduce symptoms, e.g. limiting walking or other activities, they may indicate having no difficulties in breathing during activity.[31] A way of circumventing this type of bias is to investigate differences between symptoms that are activity-dependent and those that are not. In our studies, the results were similar for these two types of symptoms.

Validated standardised questionnaires that are frequently used to determine and quantify degree of respiratory symptoms in patients with COPD in clinical practice include the modified Medical Research Council Dyspnoea Scale (mMRC), COPD assessment test (CAT), and St. Georges Respiratory Questionnaire (SGRQ).[14,32-34] The mMRC quantifies severity of dyspnoea through four questions, whereas CAT quantifies severity of cough, sputum production, chest pain and/or tightness, and dyspnoea through ten questions. According to the GOLD recommendations, mMRC

and CAT should be used for the assessment of symptoms in individuals with COPD for determining disease severity and treatment guidance.[14] The SGRQ is more complex and includes approximately 50 questions on respiratory symptoms and their impact on quality of life. In the CGPS, mMRC was introduced at cohort initiation, and CAT was introduced at the follow-up examination approximately 10 years later. The SGRQ was not implemented as it is very time-consuming and therefore not feasible to use in large-scale population-based cohorts, which simply will overshadow other more relevant questions. There will also be a potential risk of a low response-rate with large questionnaires.

Although these validated standardised questionnaires on respiratory symptoms are readily available in clinical practice, they are unfortunately not very often used outside the field of respiratory medicine. If, however, they somehow are used, it will not involve those with suspected or undiagnosed COPD, e.g. individuals with smoking exposure complaining about dyspnoea, cough, and phlegm. Importantly, we also included many unvalidated and unspecific respiratory symptoms that are frequently encountered in the general population to make it more representative.

Smoking exposure

Information on smoking exposure in the CGPS was obtained from the questionnaire that included a comprehensive string of questions on tobacco smoking, confirmed at the day of attendance by a healthcare professional. All participants were asked whether they were current or former smokers, age at smoking initiation and cessation, duration of smoking period, and current or former average amount of consumed tobacco in the form of number of daily consumed cigarettes with/without filters, cheroots, and cigars, and grams of weekly consumed pipe tobacco. Duration of smoking period (for current smokers until baseline examination) and current or former average amount of

consumed tobacco was used to calculate cumulative tobacco consumption in pack-years (one cheroot=three grams of tobacco, one cigar=five grams of tobacco, and one gram of tobacco=one cigarette): a pack-year was defined as 20 cigarettes or equivalent smoked daily for a year. Never-smokers were individuals who reported that they had never smoked in their entire life.

Recall bias should be considered when age at smoking initiation and cessation needs to be determined, which will inevitably affect the duration of smoking period and hence calculated cumulative tobacco consumption. It may be too simple to quantify all years of smoking exposure through an average amount of tobacco consumption. While most individuals will follow a regular routine in their smoking behaviour, some will have periods with more and less consumption, especially during periods with intermittent or persistent smoking cessation. Also, there may be differences in how cigarettes are consumed, e.g. while some will smoke halfway before the cigarette butts are thrown, others will smoke all the way down to the filter. Differences also include presence and degree of inhalation while smoking. Although this simple approach may be prone to recall bias and do not estimate smoking exposure more accurately, it is often used in the clinic and accepted by most epidemiologists. It is important to note that we cannot determine the true smoking exposure but only estimate and account for it in our analyses.

Separation of former from current smokers may also be a challenge, especially for individuals in a smoking cessation phase. Risk of smoking relapse will be highest for individuals with recent smoking cessation up to a year and gradually decrease over time. Another consideration is the group known as occasional smokers, which can often be divided into social smokers, binge smokers, and low-level smokers. These individuals will often not consider themselves as smokers, as a smoker in their belief is characterised by a daily tobacco consumption. Also, some former smokers will probably characterise themselves as never-smokers due to occasional smoking, lower tobacco

consumption, and/or a smoking history of a very short duration often as an adolescent. Usually, individuals with tobacco consumption less than 100 cigarettes in a lifespan are classified as never-smokers. Although there will be a form of misclassification with regards to smoking status, we have intentionally mostly relied on cumulative tobacco consumption to estimate life-time risk of COPD. Since detailed questions were used to estimate average amount of tobacco consumption through different types of tobacco, we also believe to have identified even individuals with very low smoking exposure, i.e. those with tobacco consumption <5 pack-years.

Current smokers could be identified more properly through measurement of nicotine or cotinine (the primary metabolite of nicotine) in blood, urine, saliva, or hair.[35] Measurement of cotinine is mostly preferable, as it has approximately ten times longer half-life than nicotine.[36] While non-smokers have estimated cotinine blood concentration of <1 ng/mL, but sometimes in the range of 1-10 ng/mL due to environmental tobacco smoke exposure, current smokers almost always have estimated cotinine blood concentration >10 ng/mL and sometimes >500 ng/mL.[37] Cotinine can be detected up to 10-14 days after the last cigarette consumption. Although this method seems optimal for identifying active daily smokers with high tobacco consumption, it will still be prone to the aforementioned misclassification of former smokers, occasional smokers, and smokers with low tobacco consumption.

It is also relevant to take interviewer bias into account, as some healthcare professionals may be influenced by the behaviour and appearance of some participants when reviewing their questionnaires. However, regular assessment of the training was performed assuring standardisation in collection of information. In addition, since multiple types of exposures are included in the comprehensive questionnaire, the healthcare professionals are often unaware of which exposures we are particularly interested in as well as which outcomes are used in our analyses. Questionnaires

were also applied and reviewed before lung function measurement, thereby ensuring complete blindness of presence or absence of airflow limitation for both the participant and healthcare professional.

Clinical outcomes

Clinically relevant outcomes for individuals with COPD are many and not all of them can be investigated in a population-based cohort. Selection of outcomes in the CGPS was based on relevance, practicality, and viability as well as the importance for the patient with COPD and for the clinician that treats patients with COPD. Burden of COPD can often be summarised by the increased morbidity and mortality.[14] While all-cause mortality is often easy to define and determine, morbidity may present some different challenges in clinical epidemiology. Morbidity measures traditionally include physician visits, emergency department visits, and hospitalisations. Future exacerbations and pneumonia were chosen as morbidity measures, as these outcomes are often included in clinical trials with COPD and can be determined and estimated in a setting with Danish nationwide health registries.[38-42] By combining the unique Central Person Registration number obtained from the national Civil Registration System with information on other Danish nationwide health registries such as the national Danish Patient Registry, Danish Causes of Death Registry, and Danish Cancer Registry, every single individual in Denmark can be followed from birth or immigration until death or emigration with elucidation of disease risk.[9,43] Ascertainment bias will therefore not be present in our investigations, since clinical outcomes were not obtained from the participants, but instead from nationwide health registries blinded to lung function and other relevant information.

Future exacerbations and pneumonias

Information on future exacerbations and pneumonias was obtained from the national Danish Patient Registry, which covers all public and private hospital visits in Denmark since its establishment in 1977 (including emergency department and outpatient visits since 1995).[44-46] All acute emergency department visits and hospitalisations due to exacerbation of obstructive lung disease (International Classification of Diseases [ICD]-10: J41-J46) and pneumonia with the mentioned primary discharge diagnoses were included. Thus, we were able to investigate time to first event as well as recurrent events. Individuals were followed from baseline examination until November 10, 2014 or April 10, 2018; follow-up time was updated in the CGPS during our investigations.

Exacerbation of COPD (ICD-10: J41-J44) was sometimes separated from exacerbation of asthma (ICD-10: J45-46) when the purpose was strictly to investigate individuals with COPD or clarify COPD prognosis. However, younger individuals may not be diagnosed with exacerbation of COPD during hospitalisation despite of having evidence of early or even clinical COPD, as this diagnosis by medical doctors is usually reserved for middle-aged and older individuals. We therefore used exacerbation of asthma to capture younger individuals with early COPD. Furthermore, since asthma may precede and contribute to the risk of developing COPD,[16] it would only be logical to also include exacerbation of asthma.

Since a publicly financed healthcare system covering both primary and secondary healthcare services is offered to all individuals living in Denmark, everybody has equal access to healthcare, thereby eliminating or reducing selection bias in the CGPS.[47] The national Danish Patient Registry contains administrative information such as date and type of hospital visit and medical information such as diagnoses and surgical procedures.[44-46] Diagnoses are reported by medical doctors according to national Danish laws using the World Health Organization (WHO) ICD-codes. Since

follow-up is done by combining the national Danish Patient Registry with the national Danish Civil Registration System through the unique Central Person Registration number provided to everyone at birth or immigration, no individual is lost to follow-up, and individuals who emigrate will be censored at the date of emigration.

All individuals in Denmark have been affiliated with a general practitioner. We only had the opportunity to investigate emergency department visits and hospitalisations and have thereby omitted general practitioner visits. While severe exacerbations and pneumonias will be captured with the present approach, mild or even moderate will not. Mild or moderate exacerbations and pneumonias will likely lead to general practitioner visits only, which typically will end up with prescription of systemic oral corticosteroid with/without antibiotics without any form of hospital visits. Only those patients that do not respond to the original treatment ordered by the general practitioner, or patients that have been assessed by the general practitioner to be a severe case that will need some form of complex intervention will be referred to the hospital. Although it would have been a further strength to include mild or moderate exacerbations and pneumonias, it could also be considered as a potential weakness. Due to certain limitations in their clinical practice, such as no or limited access to venous blood sampling, arterial blood gas analysis, or chest radiography/computed tomography (CT) like in a hospital setting, general practitioners very often rely on their clinical assessment alone and are forced to make a quick diagnosis. In this setting, a certain proportion of diagnostic errors or misclassifications are to be expected on a busy daily basis. By restricting to very severe exacerbations and pneumonias, as we did by only investigating emergency department visits and hospitalisations, where common necessary clinical information should be present to sustain a high diagnostic accuracy, we will be reducing degree of misclassification. Misclassification, however, cannot be eliminated completely, as we are operating with clinical diagnoses and must accept the premise that there will be individual variations by

medical doctors in their assessment of patients. To reduce such misclassifications, we have deliberately only chosen primary discharge diagnoses. All medical doctors have access to all the necessary clinical information of the entire hospitalisation during discharge and therefore have the optimal terms to determine and confirm the diagnosis.

It may sometimes be a challenge to differentiate between exacerbation and pneumonia. A pneumonia may cause a patient to exacerbate and should therefore be chosen as the underlying cause for hospitalisation.[48] However, the pneumonia diagnosis can sometimes not be verified due to an obscured chest radiography or a negative microbiology.[49] Inflammatory biomarkers in the blood may be increased during an exacerbation with or without presence of pneumonia, which complicates it even further.[50-52] A chest CT may help in clarifying the diagnosis. All necessary tests can be ordered by the medical doctors in order to clarify the diagnosis, since all individuals in Denmark have equal access to the healthcare system free of charge. By only including the primary discharge diagnosis and not the secondary as well, we were able to investigate exacerbations and pneumonias separately in our analyses both as time to first event as well as recurrent events, where results were similar. Although it would be suspected that individuals with a previous COPD diagnosis may be more likely to receive an exacerbation diagnosis instead of pneumonia, the majority of individuals with COPD are undiagnosed, as reported in the CGPS.[2] Thus, we believe that misclassification due to a previous COPD diagnosis is unlikely, and if present, it likely will be non-differential biasing towards the null-hypothesis as lung function measurement and other exposure variables are collected blinded to morbidity diagnoses.

Mortality

Information on vital status was obtained from the national Danish Civil Registration System, which is 100% complete and contains date of death for all residents in Denmark.[7-9] Information on cause of death was obtained from the national Danish Causes of Death Registry, which contains main and contributory causes of death for all residents in Denmark since its establishment in 1875.[53] Prior to 2007, a limited number of specially trained coders under the supervision of medical doctors of the Danish National Board of Health coded the national Danish Causes of Death Registry based on medical information from death certificates in accordance to WHO rules and ICD-codes. Since 2007, only medical doctors that have verified the death and issued the death certificates with indication of main and contributory causes of death have coded the national Danish Causes of Death Registry. Death due to respiratory disease or respiratory mortality (ICD-10: J00-J99) was based on the main cause of death. Individuals in the CGPS were followed from baseline examination until November 14, 2014 or April 19, 2018 for all-cause mortality, and until January 1, 2013 or December 31, 2016 for respiratory mortality. Since the national Danish Causes of Death Registry lags the national Danish Civil Registration System by approximately one year due to certain limitations in the system, not all deaths could be classified by cause.

All-cause mortality is a very precise and validated clinical outcome in Denmark. All individuals living in Denmark are registered in the national Danish Civil Registration System with their unique Central Person Registration number, and there will be no form of misclassification nor losses to follow-up when all-cause mortality is determined. In contrast, some misclassification would be expected with cause-specific mortality, as individual variations may arise despite of presence of similar medical information, when coded by different doctors. This may be complicated further by an autopsy-rate of <5% in Denmark.[54] To reduce degree of misclassification, we only used main

underlying cause. We chose death due to respiratory disease, as COPD has been shown to be underreported on death certificates in the national Danish Causes of Death Registry.[55] In fact, COPD as a main cause of death seems to have increased for the last two decades not only in Denmark but worldwide, which is believed to be due to improved diagnostics and special focus on COPD, especially with the establishment of GOLD in 1997 in collaboration with the National Heart, Lung, and Blood Institute (NHLBI), National Institutes of Health (NIH), and WHO.[56] However, despite of being one of the most frequent causes of death, respiratory disease in general is underreported compared to cardiovascular disease and cancer.[53,55] Therefore, we expected that respiratory mortality had a high specificity but low sensitivity. This meant that when risk estimates were significant, we were certain about the findings, but when risk estimates were non-significant, we could not completely rule-out a potential finding. Results showed that respiratory mortality had comparable risk estimates as those for exacerbation, pneumonia, and all-cause mortality but sometimes did not reach statistical significance, especially in additional subgroup- and sensitivity analyses, where we were challenged by statistical power.

Covariates

Information on covariates was obtained from different sources including the questionnaire, physical examination, biochemical analyses, and nationwide Danish health registries. Covariates were typically used to determine characteristics of clinical groups and/or were considered as additional risk factors and included in multivariable adjusted analyses as potential confounders. Some covariates had importance in additional subgroup- and sensitivity analyses.

Date of birth and sex were obtained from the national Danish Civil Registration System, which should be precise and not prone to bias. Information on different lifestyle-related risk factors were

available, including body mass index (BMI), blood pressure, alcohol consumption, and physical activity. BMI was calculated as measured weight divided by measured height squared (kg/m^2). Systolic and diastolic blood pressures were measured using automated equipment. Alcohol consumption included all different forms of alcoholic beverages reported in units per week and converted to grams (1 unit = 12 g of alcohol). Physical activity was reported as hours per week and degree of activity during leisure-time. Socioeconomic status was based on level of education, reported as years attending school or longest acquired education after school, and income, reported as annual household income. Blood biochemistry was measured using standard hospital assays and included among others blood leukocytes, blood neutrophils, blood eosinophils, plasma high-sensitive C-reactive protein, plasma fibrinogen, plasma α1-antitrypsin, plasma cholesterol, and plasma glucose. Fever or infection within the past 4 weeks was if individuals reported fever, bronchitis, or urinary tract infections up to 4 weeks before the day of enrolment. Episodes of acute bronchitis and/or pneumonias in the last 10 years, and general practitioner visits in the past 12 months were self-reported. Familial predisposition for asthma was at least one first degree relative with asthma, i.e. father, mother, and/or sibling. Childhood asthma, hay fever, or eczema was self-reported as a single question. Allergy was asthma, hay fever, and/or eczema as a reaction to food, medication, grass, flower, animal hair, and/or other allergens, reported in the questionnaire.

Misclassification due to measurement errors may arise in different steps of the physical examination and blood sampling and biochemical analyses, which will usually be non-differential biasing towards the null-hypothesis. However, standardisation, validation, and routine assessment of healthcare professionals training as well as routine calibration of equipment in the CGPS have likely minimised measurement errors. Biochemical analyses were also subjected to daily precision testing by using internal quality control material and monthly accuracy testing by using an external quality control programme to avoid or reduce measurement errors. Nonetheless, misclassification

due to recall bias should be expected when using self-reported information from the questionnaire. It may sometimes be difficult for individuals to remember, such as numbers, durations, or exposures. However, since all participants in the CGPS were subjected to the same questionnaire without any form of selection criteria, recall bias in such instances will apply to all participants and therefore be randomly distributed due to the random sampling. Individuals in the CGPS also completed the questionnaire before onset of clinical outcomes and were unaware of the purpose of our investigations, we therefore believe that misclassification has likely been non-differential and should only bias towards the null-hypothesis for later development of morbidity or mortality.

Comorbidities included cardiovascular disease, diabetes, and cancer. Information on cardiovascular disease was obtained from the national Danish Patient Registry and included inpatient and outpatient hospital visits due to ischaemic heart disease (ICD-8: 410-414 and ICD-10: I20-I25), stroke (ICD-8: 432-435 and ICD-10: I60, I61, I63-I64, G45), heart failure (ICD-8: 427.09-427.11 and ICD-10: I50), and atrial fibrillation (ICD-8: 427.93-427.94 and ICD-10: I48). Denmark used the ICD-8 until January 1, 1994 and proceeded directly to ICD-10 hereafter. Information on diabetes was based on self-report including use of anti-diabetic medication, non-fasting plasma glucose >11 mmol/L, and/or inpatient and outpatient hospital visits from the national Danish Patient Registry (ICD-8: 249-250 and ICD-10: E10-E14). Information on cancer was obtained from the national Danish Cancer Registry, which records all cancer forms in Denmark since 1943.[57,58] Cancer included both a history of cancer and active cancer (ICD-7: 140-205 and ICD-10: C00-D09). Non-melanoma skin cancers were excluded, as these are frequent and likely have a distinct aetiology from other cancer forms. Cancer diagnoses in the national Danish Cancer Registry are reported by medical doctors, require pathological confirmation, and are categorised based on location and histological examination by a trained pathologist using the WHO ICD-codes according to national Danish laws.

Cardiovascular disease identification in the national Danish Patient Registry has shown high validity before.[46] The national Danish Cancer Registry includes all cancer forms ever diagnosed in Denmark fully validated with relevant pathology.[57,58] Misclassification due to cardiovascular disease and cancer should therefore be of minor importance. However, since diabetes may be less represented in the national Danish Patient Registry, we had to identify individuals with diabetes from different sources thereby reducing degree of misclassification. Comorbidities were primarily used as potential confounders in our investigations.

Occupational exposure was reported as longer periods of exposure to dust/fumes during working life. Environmental tobacco smoking was reported as exposure to passive smoking in hours per day. Although occupational and environmental exposures could be identified more systematically by using validated questionnaires to reduce recall and other sources of bias, the covariates were mostly included due to convenience of being available. Detailed questionnaires on occupational and environmental exposure would overshadow other more relevant questions and not be feasible to apply in large-scale population-based cohorts and would also increase risk of a low response-rate.

Treatment with airway medication included any kind of medication for asthma and/or bronchitis daily or almost daily (including sprays and dry powder inhalers). Information on treatment could have been obtained from different more reliable sources such as the national Danish Registry of Medicinal Product Statistics, which records all prescriptions dispensed in pharmacies in Denmark since 1994.[59,60] Thus, it would be possible to determine type of airway medication, e.g. whether it is a short-or long-acting beta-2-agonist or inhaled corticosteroid. Although it would have been an advantage to use the national Danish Registry of Medicinal Product Statistics as a source, we would only have been able to determine treatment adherence but not compliance, that is, patients may dispense their prescription regularly but perhaps not take their medication on schedule as

prescribed. In our investigations, we used lack of treatment with airway medication in the identification of undiagnosed individuals with COPD, which may have introduced some misclassification that could have been avoided by using prescriptions for COPD medication instead. Nonetheless, previous hospital visits with COPD were also used in order to ensure a correct diagnostic status. Since we observed a similar proportion of underdiagnosed individuals with COPD as in a large international survey,[61] we believe to have estimated prevalence and prognosis of undiagnosed COPD well.

Asthma had a value of being a characteristic, risk factor, and potential confounder and is therefore also discussed in this section. Asthma was based on self-report and/or inpatient and outpatient hospital visits from the national Danish Patient Registry (ICD-8: 493 and ICD-10: J45-J46). Since most patients with asthma are often followed in general practice, we therefore relied mostly on self-reported information. In order to capture the most severe cases, we have also used previous hospital contacts due to asthma. Although a more clinically valid way of identifying individuals with asthma would be preferred such as presence of reversible/variable lung function or airflow limitation and/or airway hyperresponsiveness as recommended by the Global Initiative for Asthma (GINA),[30] this is not feasible in large-scale population-based cohorts such as the CGPS. However, self-reported asthma has been shown to have high specificity and sensitivity in identification of individuals with asthma in population-based cohorts.[62,63] Individuals with asthma also had to report duration of asthma, which we used to validate their diagnosis further. Nonetheless, we still must be aware of circumstances, where some individuals will identify themselves wrongfully with or without asthma. Some individuals despite absence of asthma may still identify themselves as such, e.g. by displaying well-known symptoms of asthma and having close relatives with asthma diagnosis complaining about the same symptoms. In contrast, some individuals, despite of being on treatment for asthma, may not identify themselves as a patient with asthma, e.g. due to distrust in the general

practitioner's asthma diagnosis or is being exposed to an empirical treatment to confirm an asthma diagnosis. Such misclassifications likely will be non-differential and bias towards the null-hypothesis.

Statistical analysis

All statistical analyses were performed using STATA/SE 13.1 for Windows (StataCorp, College Station, Texas, US). A two-sided P-value <0.05 was chosen to indicate statistical significance.

Group comparison

All clinical group comparisons of characteristics, risk factors, and/or potential confounders at baseline and final examination were investigated using Wilcoxon's rank-sum (alternatively Mann-Whitney), Pearson's chi-squared, and Fisher's exact tests as appropriate.[64,65] Wilcoxon's rank-sum test is non-parametric that was used to compare a continuous variable of two independent groups without having to assume a normal distribution. Pearson's chi-squared test was used to compare frequency of two independent groups, but Fisher's exact test was used in situations with small frequencies, where the number in each cell was <5 individuals.[64] However, statistically significance did often not change when using Fischer's exact instead of Pearson's chi-squared test. The downside to using non-parametric tests is that differences between groups were not quantified; however, summary data was presented for inspection alongside P-values.

Linear- and logistic regression

Since age differed substantially between some of the clinical groups during our investigations, we had to use multiple linear- and logistic regression models to account for it in the comparison of characteristics, risk factors, and/or potential confounders at the baseline examination, which would not be possible using Wilcoxon's rank-sum, Pearson's chi-squared, or Fisher's exact tests. While multiple linear regression models were additionally used to obtain reference equations for FEV_1, FVC, and FEV_1/FVC in the CGPS, so the predicted values and LLN could be calculated, logistic regression models were additionally used to determine risk of clinical COPD from baseline to final examination.

A multiple linear regression model assesses the association between a continuous dependent variable against continuous and categorical independent variables by a linear function, where the least-squares method is used to fit the best model to the observed data.[64-66] Assumptions include normal distribution of residuals, constant variance (homoscedasticity), and linearity between dependent and independent variables. Assumptions were investigated by visual inspections of scatter plots of dependent versus independent variables and its residuals and inclusion of quadratic terms to test for non-linearity. Logarithmic transformation was sometimes used in order to obtain a normal distribution.

A multiple logistic regression model uses a similar approach as a multiple linear regression model, but the dependent variable is dichotomous and assumptions such as normal distribution or homoscedasticity are not required; however, continuous independent variables still need to display linearity on a log-scale.[64-66] Another assumption is also complete follow-up time and negligibility of time-to-event as it only uses information on cumulative incident during a fixed time-period.

However, these assumptions are often not met, which often requires use of other sophisticated models better suited for accounting for time-to-event such as survival analysis.

Survival analysis

Risk of future exacerbations, pneumonias, and mortality was investigated using survival analysis with a Kaplan-Meier estimator and Cox proportional hazard function.

Kaplan-Meier estimator determines probability of survival or failure (1-survival function) in a given length of time.[64-66] Survival or failure probability was displayed graphically against analysis time, and group differences or trends was assessed with a log-rank test. While we were able to determine cumulative incidence for all-cause mortality with this approach, we were not be able to determine cumulative incidence for exacerbations or pneumonias due to competing events and risks. Age was used as analysis time and hence underlying timescale. Although it is usually recommended to use study entry as a timescale when exposures can be defined at a certain time-point such as at baseline examination, e.g. in randomised controlled trials, age is a more appropriate timescale in population-based cohorts, as it is a timescale associated with the largest changes in risk. Time of entry does not necessarily define any clinically or biologically important event. Choosing age as an underlying timescale will automatically include age as an adjustment and will also account for delayed time-entry at study examination (left truncation).[12]

Cox proportional hazard function uses semiparametric estimation and determines risk of an outcome by taking time-to-event into account and displays the risk estimates as hazard ratios.[64-66] No assumption is made on the value or shape of the baseline hazard (no intercept as opposed to a logistic regression model), but independent variables need to display a linear relationship with

outcome on a log-scale (similar to a logistic regression model). Another assumption is that the risk estimates need to be constant over the observation period, known as the proportionality assumption, i.e. the risk estimate for exposed individuals has to equal the risk estimate for unexposed individuals (also known as the baseline risk estimate) multiplied by a constant factor at any given time-point during follow-up. Such an assumption is needed in order to determine one risk estimate for the whole observation period. A violation of this assumption may imply an interaction between time and event (effect modification) and hence time-varying risk.[12] Proportionality assumption was investigated by visual inspections of log-log plots, Schoenfeld residuals, and Kaplan-Meier observed survival curves versus Cox predicted curves, and we did not observe any major violations during our investigations.[64-66]

Single- and multiple-failure time analyses were used with the extended Cox proportional hazard function. Individuals with COPD typically experience multiple events of exacerbation and pneumonia. Thus, after the occurrence of an event, these individuals will be at risk of the same event later on. Multiple-failure time analysis can therefore be used in order to estimate risk of recurrent events. To avoid counting a single event of exacerbation and pneumonia multiple times, we chose that hospitalised individuals during follow-up had to be clinically stable for at least 4 weeks after discharge before they could be considered at risk for a subsequent event. Among the different methods of multiple failure-time analysis, we chose the Andersen-Gill approach.[67] An important assumption of Andersen-Gill is that all failures are equal or indistinguishable, and it does not allow more than one event to occur at a given time-point. As a sensitivity analysis, we always carried out single-failure time analyses to investigate differences in risk estimates, and results were often similar. Marginal mean/rate approach was also used and showed similar results.[68] When Andersen-Gill approach does not contain time-dependent covariates, as it did during our

investigations, risk estimates are similar with the marginal means/rates approach.[68] Single failure-time analysis was used to investigate risk of mortality, as an individual can only die once.

Competing risk

Risk of future exacerbations and pneumonias were investigated in a setting with competing risks, where all-cause mortality and emigration were considered as competing events. Competing events are seldomly uncorrelated with the outcome of interest, which is also the reason for not treating them as usual censorings.[65,69] Competing risk analysis was used to estimate risk and cumulative incidence of exacerbations and pneumonias according to the Fine-Gray approach.[70] Fine-Gray uses semiparametric estimation that is a direct analogue to Cox proportional hazard function, but the risk estimate is displayed as subdistribution hazard ratios (subhazard ratios) in the presence of competing events. It is important to note that a subhazard hazard ratio is not the same as a hazard ratio and can therefore not be considered equivalent. Risk estimates denote the relative change in the rate of the occurrence of an outcome in individuals that have not yet experienced the outcome of interest but may instead have experienced a competing event. Since multiple failure-time analysis does not work with Fine-Gray, we instead used single failure-time analysis. An advantage with Fine-Gray is that risk and cumulative incidence can be determined with or without adjustment for potential confounders. Nonetheless, cumulative incidence was also determined according to the Coviello-Boggess approach, a form of nonparametric estimation, and differences were assessed using Pepe-Mori.[71,72] Coviello-Boggess is analogue to Kaplan-Meier but in a setting with competing events.

Predictive value and capability

Prediction of disease or prognosis has always been regarded as an essential part in medical research. Different available statistical approaches were used to investigate predictive value and capability during our investigations, depending on the purpose and study design.

Accuracy for clinical COPD at final examination was determined by estimating sensitivity, specificity, and positive- and negative predictive values for individuals with and without early COPD at the baseline examination.[64,73,74] Hereafter, the predictive capability and discriminative accuracy were investigated by determining area under the curve (AUC) for the receiver operating characteristics (ROC).[75] While sensitivity (true positive rate) and specificity (true negative rate) can be used for rule out and rule in algorithms in clinical practice, proper usefulness can only be determined through predictive values, which are dependent on the prevalence of the disease in a given population; higher prevalence will increase the positive predictive value, i.e. likelihood of presence of clinical COPD when test is positive, but instead lower the negative predictive value, i.e. likelihood of absence of clinical COPD when the test is negative. A compromise should always be considered between sensitivity and positive predictive value versus specificity and negative predictive value.

Harrell's C statistic was instead used to determine the predictive capability and discriminative accuracy for future exacerbations and mortality during follow-up, as this is a goodness-of-fit measure for predictive models with time-to-event or censoring.[76,77]

It is important to note that our analyses for predictive capability and discriminative accuracy were not complete as many different measures to investigate performance of prediction models could be used such as net reclassification index, integrated discrimination index, and decision curve

analysis.[78] Furthermore, it cannot be used in a clinical setting, as this will require additional testing and validation in different populations with associated calibration.

Confounding

Confounding is when a spurious association between the dependent and the independent variable arises due to a third variable, the confounder, that is associated with both the dependent and the independent variables. Potential confounders were selected *a priori* as those that may confound an association between COPD-related exposure variables and clinical outcomes based on previous literature and clinical and epidemiological knowledge. Confounding was assessed in three different ways: (i) by investigating the association between potential confounders with early undiagnosed COPD (or its components lung function, symptoms, and smoking exposure) and clinical outcomes, i.e. future exacerbation, pneumonia, and mortality, (ii) by investigating the association between early undiagnosed COPD and clinical outcomes in stratified analyses with potential confounders, and (iii) by investigation the association between early undiagnosed COPD and clinical outcomes before and after adjustment for potential confounders.[12] It is important to note that with these approaches, we can only account for known and measured but not unknown or unmeasured confounders.

Consequences upon adjustment with confounders may be overestimation or exaggeration of the association (thereby having a positive effect), underestimation or attenuation (negative effect), or change the direction of the association (qualitative effect).[12] It is therefore crucial to compare risk estimates from crude analyses (unadjusted or only age and sex adjusted) with those from multivariable adjusted analyses. Large discrepancies could reveal some of these effects. Furthermore, this approach would also reveal potential residual confounding and over-adjustment.

Residual confounding may arise due to improper definition of a confounder, the confounder is not a suitable surrogate for what is intended to adjust for, omission of confounders, and/or misclassification within confounders. Over-adjustment typically arises due to the confounder being in the causal pathway and is a mediator instead, or the confounder is very strongly associated with exposure or outcome of interest. During our investigations, we did not observe large discrepancies between risk estimates from crude and multivariable adjusted analyses suggesting less importance of residual confounding and over-adjustment.

Interaction

Interaction (effect modification) was assessed according to Wald's or likelihood-ratio test and in additional subgroup- and sensitivity analyses. Interaction was assessed in two different ways: (i) by investigating the association between early undiagnosed COPD and clinical outcomes stratified on a potential effect modifier (homogeneity), and (ii) by investigating presence of discrepancy between observed and expected combined effects of the effect modifier with early undiagnosed COPD in the association with a clinical outcome.[12] It is important to note the difference between confounding and interaction. Confounding creates a non-existing association, whereas interaction modifies a true association. An effect modifier may have a synergistic or an antagonistic effect on the association between exposure and outcome. While a synergistic effect strengthens an association, an antagonistic effect diminishes or eliminates an association. During our investigations, interaction analyses were carried out as part of additional subgroup- and sensitivity analyses but also from *a priori* hypotheses based upon biological plausibility to prevent heterogeneity due to random variation, confounding, or bias. Interaction was also sometimes assessed based on requests during peer review processes.

Missing values

Since we had >99% completeness in obtained information, missing values have fortunately not been a major issue during our investigations, but we still accounted for missing values in order to obtain 100% completeness. Overall, missing values can be divided into: (i) missing completely at random (MCAR), (ii) missing at random (MAR), and (iii) missing not at random (MNAR).[66] Missing values in MCAR are independent of measured and unmeasured variables and assumed to be occurring completely by chance. Missing values in MAR is dependent on or can be accounted by measured variables. Missing values in MNAR is dependent on unmeasured variables. While we cannot account for MNAR, as we do not have access to unmeasured variables, we can certainly account for MCAR and MAR through various types of analytical approaches. It is important to note that since we do not have access to unmeasured variables, we can basically not determine type of missing values and therefore need to assume that we have MCAR and/or MAR for measured variables. Since we assumed some sort of randomness in the missing values, we chose to perform multivariate imputation using chained equations to fill out missing values; however, results were similar without the use of imputation, suggesting that missing values have less importance as a source of bias during our investigations.

References

1. Çolak Y, Afzal S, Nordestgaard BG, Lange P, Vestbo J. Importance of Early COPD In Young Adults for Development of Clinical COPD: Findings from the Copenhagen General Population Study. *Am J Respir Crit Care Med* 2020; **doi:10.1164/rccm.202003-0532OC**.

2. Çolak Y, Afzal S, Nordestgaard BG, Vestbo J, Lange P. Prognosis of asymptomatic and symptomatic, undiagnosed COPD in the general population in Denmark: a prospective cohort study. *Lancet Respir Med* 2017; **5**: 426-34.

3. Çolak Y, Afzal S, Nordestgaard BG, Vestbo J, Lange P. Young and middle-aged adults with airflow limitation according to lower limit of normal but not fixed ratio have high morbidity and poor survival: a population-based prospective cohort study. *Eur Respir J* 2018; **51**: 1702681.

4. Çolak Y, Afzal S, Nordestgaard BG, Vestbo J, Lange P. Prevalence, Characteristics, and Prognosis of Early Chronic Obstructive Pulmonary Disease. The Copenhagen General Population Study. *Am J Respir Crit Care Med* 2020; **201**: 671-80.

5. Çolak Y, Nordestgaard BG, Vestbo J, Lange P, Afzal S. Prognostic significance of chronic respiratory symptoms in individuals with normal spirometry. *Eur Respir J* 2019; **54**: 1900734.

6. Çolak Y, Nordestgaard BG, Vestbo J, Lange P, Afzal S. Comparison of five major airflow limitation criteria to identify high-risk individuals with COPD: a contemporary population-based cohort. *Thorax* 2020; **75**: 944-54.

7. Pedersen CB. The Danish Civil Registration System. *Scand J Public Health* 2011; **39**: 22-5.

8. Pedersen CB, Gotzsche H, Moller JO, Mortensen PB. The Danish Civil Registration System. A cohort of eight million persons. *Dan Med Bull* 2006; **53**: 441-9.

9. Schmidt M, Pedersen L, Sorensen HT. The Danish Civil Registration System as a tool in epidemiology. *Eur J Epidemiol* 2014; **29**: 541-9.

10. Çolak Y, Afzal S, Nordestgaard BG, Marott JL, Lange P. Combined value of exhaled nitric oxide and blood eosinophils in chronic airway disease: the Copenhagen General Population Study. *Eur Respir J* 2018; **52**.

11. Marott JL, Ingebrigtsen TS, Çolak Y, Vestbo J, Lange P. Lung Function Trajectories Leading to Chronic Obstructive Pulmonary Disease as Predictors of Exacerbations and Mortality. *Am J Respir Crit Care Med* 2020; **202**: 210-8.

12. Szklo M, Nieto FJ. Epidemiology: Beyond the Basics. 3rd ed. Burlington: Jones & Bartlett Learning; 2014.

13. Bakke PS, Rönmark E, Eagan T, et al. Recommendations for epidemiological studies on COPD. *Eur Respir J* 2011; **38**: 1261-77.

14. Vogelmeier CF, Criner GJ, Martinez FJ, 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**: 557-82.

15. Salvi SS, Barnes PJ. Chronic obstructive pulmonary disease in non-smokers. *Lancet* 2009; **374**: 733-43.

16. Martinez FJ, Han MK, Allinson JP, et al. At the Root: Defining and Halting Progression of Early Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* 2018; **197**: 1540-51.

17. Soriano JB, Polverino F, Cosio BG. What is early COPD and why is it important? *Eur Respir J* 2018; **52**: 1801448.

18. Løkke A, Marott JL, Mortensen J, Nordestgaard BG, Dahl M, Lange P. New Danish reference values for spirometry. *Clin Respir J* 2013; **7**: 153-67.

19. Bellamy D. Spirometry in practice: a practical guide to using spirometry in primary care. 2nd ed. London: The BTS COPD Consortium; 2005.

20. Bellia V, Pistelli R, Catalano F, et al. Quality control of spirometry in the elderly. The SA.R.A. study. SAlute Respiration nell'Anziano = Respiratory Health in the Elderly. *Am J Respir Crit Care Med* 2000; **161**: 1094-100.

21. Pezzoli L, Giardini G, Consonni S, et al. Quality of spirometric performance in older people. *Age Ageing* 2003; **32**: 43-6.

22. Çolak 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**: 172-81.

23. Quanjer PH, Stanojevic S, Cole TJ, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. *Eur Respir J* 2012; **40**: 1324-43.

24. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med* 1999; **159**: 179-87.

25. Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF, Peslin R, Yernault JC. Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. *Eur Respir J Suppl* 1993; **16**: 5-40.

26. Lange P, Çolak Y, Ingebrigtsen TS, Vestbo J, Marott JL. Long-term prognosis of asthma, chronic obstructive pulmonary disease, and asthma-chronic obstructive pulmonary disease overlap in the Copenhagen City Heart study: a prospective population-based analysis. *Lancet Respir Med* 2016; **4**: 454-62.

27. Hansen JG, Pedersen L, Overvad K, Omland O, Jensen HK, Sorensen HT. The Prevalence of chronic obstructive pulmonary disease among Danes aged 45-84 years: population-based study. *COPD* 2008; **5**: 347-52.

28. Mohamed Hoesein FA, Zanen P, Sachs AP, Verheij TJ, Lammers JW, Broekhuizen BD. Spirometric thresholds for diagnosing COPD: 0.70 or LLN, pre- or post-dilator values? *COPD* 2012; **9**: 338-43.

29. Thomsen RW, Lange P, Hellquist B, et al. Validity and underrecording of diagnosis of COPD in the Danish National Patient Registry. *Respir Med* 2011; **105**: 1063-8.

30. Global Initiative for Asthma. Global Strategy for Asthma Management and Prevention 2020. Available from: www.ginasthma.com (Last accessed December 22, 2020).

31. Miravitlles M, Ribera A. Understanding the impact of symptoms on the burden of COPD. *Respir Res* 2017; **18**: 67.

32. Bestall JC, Paul EA, Garrod R, Garnham R, Jones PW, Wedzicha JA. Usefulness of the Medical Research Council (MRC) dyspnoea scale as a measure of disability in patients with chronic obstructive pulmonary disease. *Thorax* 1999; **54**: 581-6.

33. Jones PW, Harding G, Berry P, Wiklund I, Chen WH, Kline Leidy N. Development and first validation of the COPD Assessment Test. *Eur Respir J* 2009; **34**: 648-54.

34. Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire. *Am Rev Respir Dis* 1992; **145**: 1321-7.

35. Raja M, Garg A, Yadav P, Jha K, Handa S. Diagnostic Methods for Detection of Cotinine Level in Tobacco Users: A Review. *J Clin Diagn Res* 2016; **10**: ZE04-6.

36. Benowitz NL, Jacob P, 3rd. Metabolism of nicotine to cotinine studied by a dual stable isotope method. *Clin Pharmacol Ther* 1994; **56**: 483-93.

37. Hukkanen J, Jacob P, 3rd, Benowitz NL. Metabolism and disposition kinetics of nicotine. *Pharmacol Rev* 2005; **57**: 79-115.

38. Cazzola M, MacNee W, Martinez FJ, et al. Outcomes for COPD pharmacological trials: from lung function to biomarkers. *Eur Respir J* 2008; **31**: 416-69.

39. Celli BR, Decramer M, Wedzicha JA, et al. An official American Thoracic Society/European Respiratory Society statement: research questions in COPD. *Eur Respir J* 2015; **45**: 879-905.

40. Glaab T, Vogelmeier C, Buhl R. Outcome measures in chronic obstructive pulmonary disease (COPD): strengths and limitations. *Respir Res* 2010; **11**: 79.

41. Jones PW, Beeh KM, Chapman KR, Decramer M, Mahler DA, Wedzicha JA. Minimal clinically important differences in pharmacological trials. *Am J Respir Crit Care Med* 2014; **189**: 250-5.

42. Mackay AJ, Kostikas K, Murray L, et al. Patient-reported Outcomes for the Detection, Quantification, and Evaluation of Chronic Obstructive Pulmonary Disease Exacerbations. *Am J Respir Crit Care Med* 2018; **198**: 730-8.

43. Erlangsen A, Fedyszyn I. Danish nationwide registers for public health and health-related research. *Scand J Public Health* 2015; **43**: 333-9.

44. Andersen TF, Madsen M, Jorgensen J, Mellemkjoer L, Olsen JH. The Danish National Hospital Register. A valuable source of data for modern health sciences. *Dan Med Bull* 1999; **46**: 263-8.

45. Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. *Scand J Public Health* 2011; **39**: 30-3.

46. Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sorensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol* 2015; **7**: 449-90.

47. Schmidt M, Schmidt SAJ, Adelborg K, et al. The Danish health care system and epidemiological research: from health care contacts to database records. *Clin Epidemiol* 2019; **11**: 563-91.

48. Wedzicha JA, Seemungal TA. COPD exacerbations: defining their cause and prevention. *Lancet* 2007; **370**: 786-96.

49. Musher DM, Thorner AR. Community-acquired pneumonia. *N Engl J Med* 2014; **371**: 1619-28.

50. Crisafulli E, Manco A, Ferrer M, et al. Pneumonic versus Nonpneumonic Exacerbations of Chronic Obstructive Pulmonary Disease. *Semin Respir Crit Care Med* 2020; **41**: 817-29.

51. Huerta A, Crisafulli E, Menendez R, et al. Pneumonic and nonpneumonic exacerbations of COPD: inflammatory response and clinical characteristics. *Chest* 2013; **144**: 1134-42.

52. Lieberman D, Lieberman D, Gelfer Y, et al. Pneumonic vs nonpneumonic acute exacerbations of COPD. *Chest* 2002; **122**: 1264-70.

53. Helweg-Larsen K. The Danish Register of Causes of Death. *Scand J Public Health* 2011; **39**: 26-9.

54. Larsen ST, Lynnerup N. Medico-legal autopsies in Denmark. *Dan Med Bull* 2011; **58**: A4247.

55. Jensen HH, Godtfredsen NS, Lange P, Vestbo J. Potential misclassification of causes of death from COPD. *Eur Respir J* 2006; **28**: 781-5.

56. Rodriguez-Roisin R, Rabe KF, Vestbo J, et al. Global Initiative for Chronic Obstructive Lung Disease (GOLD) 20th Anniversary: a brief history of time. *Eur Respir J* 2017; **50**.

57. Gjerstorff ML. The Danish Cancer Registry. *Scand J Public Health* 2011; **39**: 42-5.

58. Storm HH, Michelsen EV, Clemmensen IH, Pihl J. The Danish Cancer Registry--history, content, quality and use. *Dan Med Bull* 1997; **44**: 535-9.

59. Kildemoes HW, Sorensen HT, Hallas J. The Danish National Prescription Registry. *Scand J Public Health* 2011; **39**: 38-41.

60. Pottegård A, Schmidt SAJ, Wallach-Kildemoes H, Sorensen HT, Hallas J, Schmidt M. Data Resource Profile: The Danish National Prescription Registry. *Int J Epidemiol* 2017; **46**: 798-f.

61. Lamprecht B, Soriano JB, Studnicka M, et al. Determinants of underdiagnosis of COPD in national and international surveys. *Chest* 2015; **148**: 971-85.

62. Peat JK, Toelle BG, Marks GB, Mellis CM. Continuing the debate about measuring asthma in population studies. *Thorax* 2001; **56**: 406-11.

63. Toren K, Brisman J, Jarvholm B. Asthma and asthma-like symptoms in adults assessed by questionnaires. A literature review. *Chest* 1993; **104**: 600-8.

64. Altman DG. Practical Statistics for Medical Research. 1st ed. London: Chapman & Hall; 1991.

65. Clayton D, Hills M. Statistical Models in Epidemiology. 1st ed. Oxford: Oxford University Press; 2013.

66. Jewell NP. Statistics for Epidemiology. London: Chapman & Hall; 2004.

67. Andersen PK, Gill RD. Cox's regression model for counting processes: A large sample study. *Annals of Statistics* 1982; **10**: 1100-20.

68. Amorim LD, Cai J. Modelling recurrent events: a tutorial for analysis in epidemiology. *Int J Epidemiol* 2015; **44**: 324-33.

69. Cleves M, Gould WW, Marchenko Y. An Introduction to Survival Analysis Using Stata. 3rd ed. Texas: Stata Press; 2010.

70. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* 1999; **94**: 496-509.

71. Coviello V, Boggess M. Cumulative incidence estimation in the presence of competing risks. *The Stata Journal* 2004; **4**: 103-12.

72. Pepe MS, Mori M. Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Stat Med* 1993; **12**: 737-51.

73. Altman DG, Bland JM. Diagnostic tests 2: Predictive values. *BMJ* 1994; **309**: 102.

74. Altman DG, Bland JM. Diagnostic tests. 1: Sensitivity and specificity. *BMJ* 1994; **308**: 1552.

75. Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. *BMJ* 1994; **309**: 188.

76. Harrell FE, Jr., Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA* 1982; **247**: 2543-6.

77. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; **15**: 361-87.

78. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010; **21**: 128-38.

Part III – Original articles

List of publications forming the doctoral thesis

1. **Çolak Y**, Afzal S, Nordestgaard BG, Vestbo J, Lange P. Prognosis of asymptomatic and symptomatic, undiagnosed COPD in the general population in Denmark: a prospective cohort study. *Lancet Respir Med*. 2017; 5(5): 426-434.

2. **Çolak Y**, Afzal S, Nordestgaard BG, Vestbo J, Lange P. Young and middle-aged adults with airflow limitation according to lower limit of normal but not fixed ratio have high morbidity and poor survival: a population-based prospective cohort study. *Eur Respir J*. 2018; 51(3): 1702681.

3. **Çolak Y**, Nordestgaard BG, Vestbo J, Lange P, Afzal S. Prognostic significance of chronic respiratory symptoms in individuals with normal spirometry. *Eur Respir J*. 2019; 54(3): 1900734.

4. **Çolak Y**, Afzal, Nordestgaard BG, Vestbo J, Lange P. Prevalence, characteristics, and prognosis of early chronic obstructive pulmonary disease. The Copenhagen General Population Study. *Am J Respir J Crit Care Med*. 2020; 201(6): 671-680.

5. **Çolak Y**, Nordestgaard BG, Vestbo J, Lange P, Afzal S. Comparison of five major airflow limitation criteria to identify high-risk individuals with COPD: a contemporary population-based cohort. *Thorax*. 2020;75(11): 944-954.

6. **Çolak Y**, Afzal S, Nordestgaard BG, Lange P, Vestbo J. Importance of early COPD in young adults for development of clinical COPD: Findings from the Copenhagen General Population Study. *Am J Respir J Crit Care Med*. 2020; DOI:10.1164/rccm.202003-0532OC.