



Factors most strongly associated with breathlessness in a population aged 50–64 years

Max Olsson ¹, Anders J. Björkelund², Jacob Sandberg ¹, Anders Blomberg ³, Mats Börjesson^{4,5}, David Currow⁶, Andrei Malinowski ⁷, Magnus Sköld^{8,9}, Per Wollmer¹⁰, Kjell Torén¹¹, Carl Johan Östgren^{12,13}, Gunnar Engström ¹⁴ and Magnus Ekström ¹

¹Lund University, Faculty of Medicine, Department of Clinical Sciences Lund, Respiratory Medicine, Allergology and Palliative Medicine, Lund, Sweden. ²Lund University, Faculty of Science, Centre for Environmental and Climate Science, Lund, Sweden. ³Umeå University, Department of Public Health and Clinical Medicine, Umeå, Sweden. ⁴Dept of Molecular and Cardiovascular Medicine, Center for Health and Performance, Sahlgrenska Academy, Gothenburg, Sweden. ⁵Dept MGAÖ, Sahlgrenska University Hospital, Gothenburg, Sweden. ⁶Faculty of Science, Medicine and Health, University of Wollongong, Wollongong, Australia. ⁷Dept of Medical Sciences, Clinical Physiology, Uppsala University, Uppsala, Sweden. ⁸Karolinska Institutet, Department of Medicine Solna, Stockholm, Sweden. ⁹Department of Respiratory Medicine and Allergy, Karolinska University Hospital, Stockholm, Sweden. ¹⁰Department of Translational Medicine, Lund University, Malmö, Sweden. ¹¹Occupational and Environmental Medicine, School of Public Health and Community Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. ¹²Centre of Medical Image Science and Visualization, Linköping University, Linköping, Sweden. ¹³Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden. ¹⁴Lund University, Faculty of Medicine, Department of Clinical Sciences Malmö, Cardiovascular Epidemiology, Malmö, Sweden.

Corresponding author: Max Olsson (max.olsson@med.lu.se)



Shareable abstract (@ERSpublications)

Increased body mass index, reduced lung function, sedentary lifestyle are the factors most strongly associated to breathlessness in the general population. These factors may be prevented or amenable to public health interventions. <https://bit.ly/3so80Ce>

Cite this article as: Olsson M, Björkelund AJ, Sandberg J, *et al.* Factors most strongly associated with breathlessness in a population aged 50–64 years. *ERJ Open Res* 2024; 10: 00582-2023 [DOI: 10.1183/23120541.00582-2023].

Copyright ©The authors 2024

This version is distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0. For commercial reproduction rights and permissions contact permissions@ersnet.org

Received: 11 Aug 2023
Accepted: 31 Oct 2023

Abstract

Background Breathlessness is a troublesome and prevalent symptom in the population, but knowledge of related factors is scarce. The aim of this study was to identify the factors most strongly associated with breathlessness in the general population and to describe the shapes of the associations between the main factors and breathlessness.

Methods A cross-sectional analysis was carried out of the multicentre population-based Swedish CARDioPulmonary bioImage Study (SCAPIS) of adults aged 50 to 64 years. Breathlessness was defined as a modified Medical Research Council breathlessness rating ≥ 2 . The machine learning algorithm extreme gradient boosting (XGBoost) was used to classify participants as either breathless or nonbreathless using 449 factors, including physiological measurements, blood samples, computed tomography cardiac and lung measurements, lifestyle, health conditions and socioeconomics. The strength of the associations between the factors and breathlessness were measured by SHapley Additive exPlanations (SHAP), with higher scores reflecting stronger associations.

Results A total of 28 730 participants (52% women) were included in the study. The strongest associated factors for breathlessness were (in order of magnitude): body mass index (SHAP score 0.39), forced expiratory volume in 1 s (0.32), physical activity measured by accelerometry (0.27), sleep apnoea (0.22), diffusing lung capacity for carbon monoxide (0.21), self-reported physical activity (0.17), chest pain when hurrying (0.17), high-sensitivity C-reactive protein (0.17), recent weight change (0.14) and cough (0.13).

Conclusion This large population-based study of men and women aged 50–64 years identified the main factors related to breathlessness that may be prevented or amenable to public health interventions.

Introduction

Breathlessness is defined as the experience of breathing discomfort [1] and is a public health issue of concern, as 10–25% of the middle-aged and older general population experience, and are limited by, that



symptom in their daily life [2, 3]. The symptom is associated with increased mortality [4], often compromising the affected individual's daily life activities [5] and quality of life [6]. Breathlessness can arise through a complex interplay of multiple factors, including physiological, psychological, social and environmental factors [5, 7, 8]. There are multiple health conditions known to contribute to breathlessness in the general population, such as cardiovascular and respiratory diseases, obesity, anxiety and depression [1, 2]. The majority of individuals experiencing breathlessness have several concurrent underlying conditions, [1] and which factors contribute to breathlessness can be difficult to untangle [1]. After evaluation at a specialist breathlessness clinic, the underlying condition(s) contributing to a patient's breathlessness may remain unknown among one-third of the patients [3].

There is a lack of studies comparing the associations of multiple factors with breathlessness at a detailed level in the general population [1]. Previous studies have chiefly used traditional statistical methods to evaluate associations for one or a limited number of conditions with breathlessness but have not considered the complex interplay and potential nonlinear associations between breathlessness and multiple concurrent factors, such as physiological, psychological, social and environmental factors [1]. When using large datasets, traditional statistical regression models are limited in their ability to evaluate multiple factors that are concurrently contributing and interplaying in complex associations to an outcome. The recent developments in data-driven methods could help us to identify the strongest associated factors related to breathlessness in the general population [1].

Machine learning has emerged as a novel method to analyse complex associations with benefits over traditional statistical methods when analysing multiple factors with complex, nonlinear associations. Machine learning can rank the strongest associated factors for an outcome and analyse the interplay between multiple factors in relation to the outcome [4].

The aims of this study were to identify the factors most strongly associated with breathlessness using a machine learning approach and to describe the shape of the association between the main factors and breathlessness in the middle-aged general population.

Methods

Study design and population

This was an analysis of the multicentre, population-based Swedish CARdioPulmonary bioImage Study (SCAPIS) of ~30 000 men and women aged 50 to 64 years [5]. The SCAPIS data were collected between 2013 and 2018 at six study sites located in Uppsala, Umeå, Linköping, Malmö/Lund, Gothenburg and Stockholm. The participants were randomly invited from the population in the areas surrounding the study sites. The participants needed to be able to understand instructions and complete questionnaires in Swedish. At the study sites, participants underwent physiological tests and completed a questionnaire about lifestyle, health conditions and other symptoms, as specified below. The SCAPIS cohort has been shown to be representative of the Swedish general population of similar ages [6]. A previous study was conducted with data from the SCAPIS study that looked at prespecified medical conditions associated with breathlessness using traditional logistic regression models [2]. The present study is significantly different from the previous one, as the present study uses a novel analysis method that can compare multiple factors' nonlinear associations, which has not been presented in this context before.

Exclusion criteria for the present analysis were missing data on breathlessness (the modified Medical Research Council (mMRC) breathlessness scale) [7] and inability to walk for reasons other than breathlessness.

Assessments

Breathlessness was defined as a self-rated mMRC score ≥ 2 ("breathless when walking on level ground" or worse) [7]. A total of 449 factors were included in the analysis, which includes all available factors in the SCAPIS study in addition to specific diet questions, as well as proteomics and metabolic data that were only available in a subsample of 5000 subjects. All factors and assessments are detailed in supplementary table S1. The evaluated factors included self-reported health conditions diagnosed by a physician, measured body mass index (BMI), spirometry (forced expiratory volume in 1 s (FEV₁) and forced vital capacity (FVC) post bronchodilation, diffusing lung capacity for carbon monoxide (D_{LCO}), computed tomography (CT) cardiac and lung measurements, blood samples (high-sensitivity C-reactive protein (hsCRP), creatinine, cholesterol, glucose, haemoglobin, HbA1c, high-density lipoprotein, low-density lipoprotein, triglycerides) and self-reported physical activity as well as that measured by accelerometry during 1 week (mean and total minutes of different intensity levels). Self-reported data included physician-diagnosed health conditions (sleep apnoea, COPD), other symptoms (coughing, wheezing in

chest), treatments, lifestyle factors and heredity of health conditions. Socioeconomic factors were both self-reported (education level, income, housing type) and derived from neighbourhood demographics (mean income, ratio of residents with foreign background, university degree) from Statistics Sweden [8]. Absolute spirometry values were used, as the other variables used to calculate % predicted values, such as age, height and sex, were also included in the analysis [9, 10].

Machine learning

Machine learning was used to evaluate which of the factors were most strongly associated with breathlessness in this general population sample. This approach has previously been validated in a pilot study of self-perceived health in older men [11].

Machine learning methods have been comprehensively reviewed elsewhere [12]. Briefly, machine learning algorithms are complex algorithms that can determine nonlinear associations between factors and an outcome, accounting for all factors and their complex interplay in the dataset. To measure the generalisability of the machine learning model, the data are split into a training and a test set. The model is first trained to classify an outcome, where the algorithm adapts the model to the training data such that it can optimally classify the outcome in the training data. The model is subsequently evaluated by measuring its outcome classification performance on the test set [12].

Statistical analyses

Patient characteristics were summarised using descriptive statistics. The Stockholm study site (n=4912; 17%) was used as the test set, and the other study sites were used as the training set. The model was trained with a fivefold cross-validation on the training set with the area under the curve (AUC) of the receiver operating characteristic (ROC) as the target. The machine learning algorithm extreme gradient boosting (XGBoost) [13] was trained to classify participants as breathless or not breathless. XGBoost is a modern and popular machine learning algorithm for the classification of outcomes using tabular data [13]. XGBoost is a gradient boosting model that uses an ensemble of decision trees to achieve better classification [13]. Missing values were not imputed; instead, the XGBoost model chose a path in the decision tree for missing values for each factor [13]. To determine the best set of hyperparameters (settings of the model that can improve performance), the model was trained with a fivefold cross-validation on the training set with the AUC of the ROC as the target. The final model was evaluated on the test set by sensitivity, specificity and AUC. When classifying breathlessness on the test set, the model had a sensitivity of 0.73, a specificity of 0.89 and an AUC of 0.81, which means that the model is able to generalise for unseen data. See supplementary table S2 for details about the final model. As a sensitivity analysis, we tested mMRC 1 as the outcome, but we were unable to train the model because it ceased to be able to discriminate between people who were breathless and those who were not breathless.

The strength of associations between factors and breathlessness was measured using SHapley Additive exPlanations (SHAP), which have emerged as a robust assessment of the strength of associations in machine learning studies [4]. Participants receive an individual SHAP score for each factor in the dataset, representing a change in the probability of being classified as breathless, which can be either positive (increased probability) or negative (decreased probability). We described two measurements of the associations between individual factors and breathlessness, the first being the strength of the association by SHAP absolute mean to identify the factors with the strongest associations with breathlessness. Second, we described the shape of the associations in scatter plots to describe how a change in a factor's value changes the probability of being breathless, meaning that the shape of the associations can be used to identify the threshold of a factor when breathlessness is most likely, adjusted for all other factors. To simplify the interpretation of the shapes of the associations in these figures, the SHAP values were first transformed to probabilities (range: 0–1), and then locally estimated scatterplot smoothing lines were added to the scatter plots. The analyses were conducted using R software (version 4.1.2; R Foundation for Statistical Computing, Austria) together with the “xgboost” package (version 1.6.0.1).

Ethical considerations

The SCAPIS multicentre study was approved by the ethics committee at Umeå University (Dnr 2010–228–31 M), and the present analysis was approved by the Swedish Ethical Review Authority (reference number: 2021-00288). All participants gave their written informed consent.

Results

Participant characteristics

A total of 14 801 (52%) women and 13 929 (48%) men were included in the analysis. Participants' mean±SD age was 57.5±4.3 and their mean BMI was 26.9±4.4; 21% were obese (BMI ≥30), 8% had

chronic airflow limitation ($FEV_1/FVC < 0.7$), 12% were regular or occasional smokers, 36% were former smokers and 50% were never-smokers (table 1). A physician diagnosis of asthma was self-reported by 8%, COPD, chronic bronchitis or emphysema by 1%, and heart failure by 1% of the participants. Breathlessness was present in 4.4% of the total sample, in 3% of the men, and in 6% of the women (table 1).

Factors most strongly associated with breathlessness

The factors most strongly associated with breathlessness were (SHAP absolute mean): BMI (0.39), FEV_1 (0.32), vigorous-intensity physical activity measured by accelerometry (0.27), self-reported sleep apnoea (0.22), D_{LCO} (0.21), self-reported physical activity (0.17), self-reported chest pain when hurrying (0.17), hsCRP (0.17), self-reported recent weight change (0.14) and self-reported cough (when not having a cold) (0.13) (figure 1). Supplementary table S3 shows the strength of the association of all factors with a SHAP absolute mean above zero.

Shapes of associations between main factors and breathlessness

The shapes of the associations for the 10 factors with the strongest associations with breathlessness are shown in figure 2. BMI had an S-shaped association with breathlessness, with a BMI > 30 being positively associated with breathlessness, and the association increased along with increased BMI. Lung function measured by FEV_1 also had an S-shaped association with breathlessness, with a value of < 3 L being associated with breathlessness, and the association increased along with decreased FEV_1 . Lung function measured by D_{LCO} had a similar association with breathlessness as with FEV_1 , with < 8 $\text{mmol}\cdot\text{min}^{-1}\cdot\text{kPa}^{-1}$ being associated with breathlessness. A sedentary lifestyle without vigorous-intensity physical activity during the week of accelerometry measurement was positively associated with breathlessness, but only a few minutes a week was needed for the association to be negatively associated with breathlessness. A similar pattern could also be seen with self-reported physical activity, as a sedentary lifestyle was associated with breathlessness. Sleep apnoea, chest pain when hurrying, increased hsCRP, coughing while not having a cold and having a recent weight change were all positively associated with breathlessness (figure 2).

Discussion

Main findings

In this large population-based study of middle-aged people, we used machine learning to identify the most strongly associated factors with breathlessness as determined by mMRC scores. Increased BMI, reduced lung function, sedentary lifestyle, sleep apnoea, chest pain while hurrying, increased hsCRP, coughing and weight changes were the factors that were most strongly related to breathlessness in the study population. These findings confirm that breathlessness arises from a multitude of factors and identify targets for potential public health interventions to decrease the burden of breathlessness in the population.

What this study adds

This is the first study with a broad and multifactorial focus on breathlessness in the general population, including physiological, psychological, social and environmental factors. The analysis was not driven by the investigators' hypotheses but instead used an explorative data-driven approach to untangle complex interplays between factors and breathlessness with as few presumptions as possible. The association between increased BMI, heart diseases, lung diseases and breathlessness is well studied [1], but our study evaluated multiple aspects beyond health conditions and quantified the strength as well as the shapes of the associations.

Increased BMI was the factor most strongly associated with breathlessness in our study. Since our analysis adjusted for all other evaluated factors in the model, including medical conditions, our findings suggest that increased BMI itself is directly associated with breathlessness and that the association is not mediated through any higher prevalence of other conditions or factors in people with higher BMI. A direct association between increased BMI and breathlessness has been presented before [10, 14]. This confirms and extends the conclusions of previous reports on the BMI–breathlessness association, which were limited by the fact that the analysis was only adjusted for lung function and a few comorbidities [10], making the evidence sparse, and our study reinforces the evidence of a direct association of increased BMI with breathlessness. The extra workload by a heavier body resulting in an increased respiratory demand could explain this association, as suggested before [15]. From a public health and clinical perspective, our findings are also valuable, as they suggest that the World Health Organisation's (WHO) classification of obesity (BMI > 30) is an adequate threshold, reflecting a mean level when BMI relates to markedly higher breathlessness affecting daily life activities [15].

TABLE 1 Characteristics of 28 730 men and women aged 50–64 years in the general population

Factor	All	With breathlessness	Without breathlessness
Sex, female	14 801 (52)	830 (6)	13 971 (94)
Sex, male	13 929 (48)	379 (3)	13 550 (97)
Age years	57.5±4.3	58.18±4.4	57.47±4.3
Born in Sweden	24 153 (84)	793 (66)	23 360 (85)
BMI kg·m⁻²	26.9±4.4	31.00±6.0	26.72±4.2
Overweight (BMI ≥25)	18 451 (64)	1040 (86)	17 411 (63)
Obese (BMI ≥30)	5994 (21)	654 (54)	5340 (19)
FEV₁ L	3.3±0.8	2.68±0.7	3.30±0.8
FVC L	4.2±1.0	3.55±0.9	4.24±1.0
FEV₁/FVC	0.8±0.1	0.76±0.1	0.78±0.1
Chronic airway obstruction (FEV₁/FVC <0.7)	2418 (8)	215 (18)	2203 (8)
Employment professional work	25 073 (87)	810 (67)	24 263 (88)
Highest completed level of education			
No elementary school	177 (1)	24 (2)	153 (1)
Elementary school	2447 (9)	210 (17)	2237 (8)
Upper secondary school	12 986 (45)	618 (51)	12 368 (45)
University degree	13 032 (45)	349 (29)	12 683 (46)
Current smoking status			
Regular	2119 (7)	169 (14)	1950 (7)
Occasional	1389 (5)	55 (5)	1334 (5)
Former	10 382 (36)	501 (41)	9881 (36)
Never	14 473 (50)	467 (39)	14 006 (51)
Pack-years of smoking	7.6±12	13.91±17	7.29±12
Frequency of having an alcoholic drink			
Never	2511 (9)	279 (23)	2232 (8)
Monthly or less	4370 (15)	290 (24)	4080 (15)
2 to 4 times a month	10 951 (38)	374 (31)	10 577 (38)
2 to 3 times a week	8653 (30)	183 (15)	8470 (31)
4 or more times a week	2079 (7)	65 (5)	2014 (7)
Self-reported exertion, last 12 months			
Sedentary	3266 (11)	420 (35)	2846 (10)
Moderate occasional exercise	13 155 (46)	558 (46)	12 597 (46)
Moderate but regular exercise	8296 (29)	124 (10)	8172 (30)
Frequent heavy exercise	3783 (13)	91 (8)	3692 (13)
Quality of sleep			
Very well	4975 (17)	116 (10)	4859 (18)
Well	8828 (31)	225 (19)	8603 (31)
Rather well	9912 (35)	387 (32)	9525 (35)
Badly	4006 (14)	352 (30)	3654 (13)
Very badly	710 (3)	111 (9)	599 (2)
Weight change	3831(13)	3360 (12)	471 (39)
Emphysema found on CT	1578 (5)	131 (11)	1447 (5)
Self-reported health conditions			
Asthma	2288 (8)	239 (20)	2049 (7)
Cancer	1698 (6)	84 (7)	1614 (6)
COPD, chronic bronchitis or emphysema	321 (1)	83 (7)	163 (1)
Diabetes	1221 (4)	114 (9)	1107 (4)
Feelings of sadness/depression	7806 (27)	634 (52)	7172 (26)
Heart failure	141 (1)	33 (3)	108 (1)
Hypertension	6407 (22)	436 (36)	5971 (22)
Myocardial infarction	449 (2)	52 (4)	397 (1)
Rheumatic disease	1006 (4)	97 (8)	909 (3)
Sleep apnoea	1240 (4)	118 (10)	1122 (4)
Stroke	394 (1)	33 (3)	361 (1)

Data are presented as factors mean±SD or n (%) within the groups “All”, “With breathlessness” and “Without breathlessness”. BMI: body mass index; FEV₁: forced expiratory volume in 1 s; FVC: forced vital capacity; CT: computed tomography.

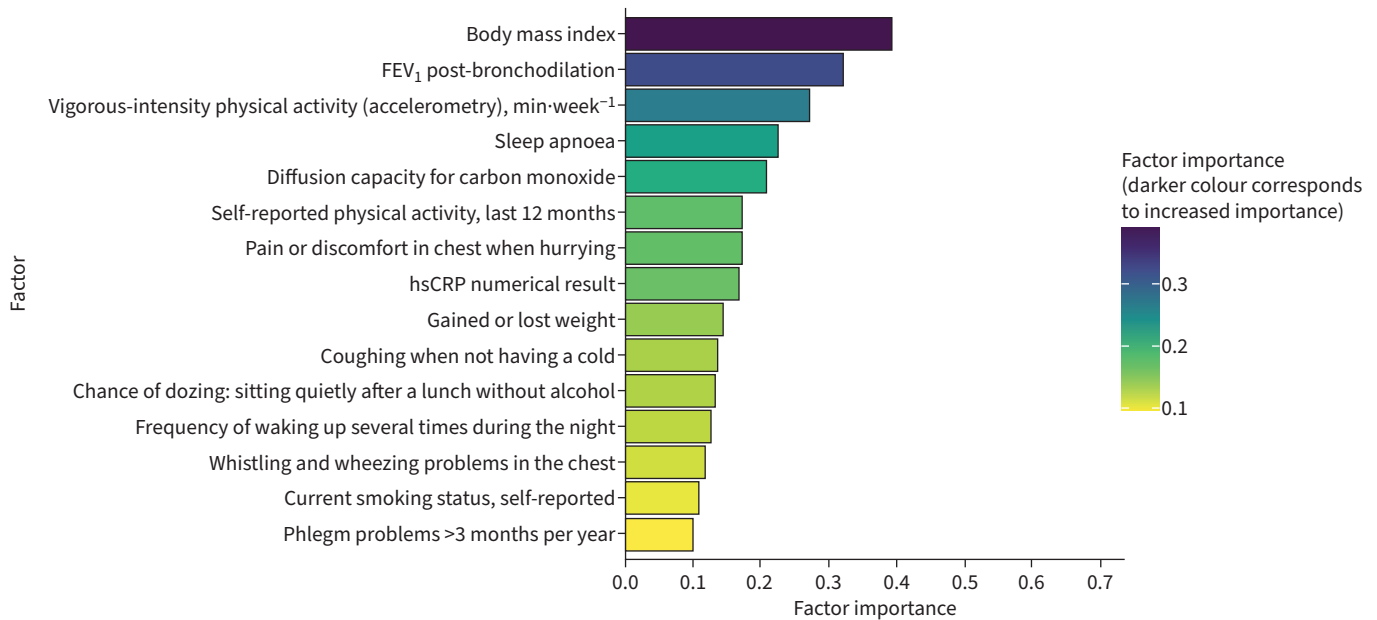


FIGURE 1 Strength of association between factor and breathlessness among 28 730 men and women aged 50–64 years in the general population. The strength of association was measured by SHapley Additive exPlanations (SHAP) absolute mean, which represents the mean contribution of a factor to the change in probability of being breathless or not breathless but does not present any direction of the association. A higher score corresponds to a stronger association. FEV₁: forced expiratory volume in 1 s; hsCRP: high-sensitivity C-reactive protein.

Physical inactivity was strongly associated with breathlessness in our study and should be seen as an additional factor associated with breathlessness independent of obesity because of the multivariate analysis. Vigorous-intensity physical activity (or exercise) measured by accelerometry was more strongly associated with breathlessness than self-reported weekly physical activity. This can be explained by the difference between reported and actual physical activity, for example, an overestimation by the participant of how much weekly physical activity he or she performs. Accelerometry is also a quantitative measurement of the intensity of physical activity, which was lacking in the self-reported measurement of physical activity used in our study. In our study, we used different accelerometry measurements, and vigorous-intensity physical activity was more strongly associated with breathlessness than low- and moderate-intensity physical activity. Vigorous-intensity physical activity might be the threshold that needs to be met to develop breathlessness, which might not be met by individuals who develop breathlessness by walking on level ground (mMRC ≥ 2). This can be related to breathlessness being known to reduce the activity of the affected individuals [16] and thereby result in a sedentary lifestyle risk, thereby trapping the affected individual in a “vicious cycle”, with breathlessness contributing to an inactive lifestyle that produces further deterioration in breathlessness [17].

Reduced lung function, measured as lower FEV₁ and/or D_{LCO} , was strongly associated with breathlessness. Interestingly, other lung function measurements, such as FVC and slow vital capacity, were not as strongly associated with breathlessness in our study. A similar shape of association between FEV₁ and breathlessness has been presented in a previous general population study, but after adjusting for confounders, no association was seen [18]. Our study adds evidence to the association between decreased FEV₁ and breathlessness, but the overall evidence is still sparse, as suggested before [1]. A novel finding is that lung function measurements were more strongly associated with breathlessness than CT measurements. The findings suggest that lung function as a physiological test more strongly correlates with the experience of exertional breathlessness compared to CT measurements. However, a study of mild COPD patients [19] showed a correlation between breathlessness and airway narrowing: a CT measurement which was not included in our study. Also, COPD was not an important factor and could be undiagnosed and incorporated within the lung function factors.

Sex was not associated with breathlessness in the present analysis. Of note, this should not be interpreted as there being no differences in breathlessness between the sexes, as 6% of the women and 3% of the men suffered from breathlessness. In line with this, women have been found to report more breathlessness on

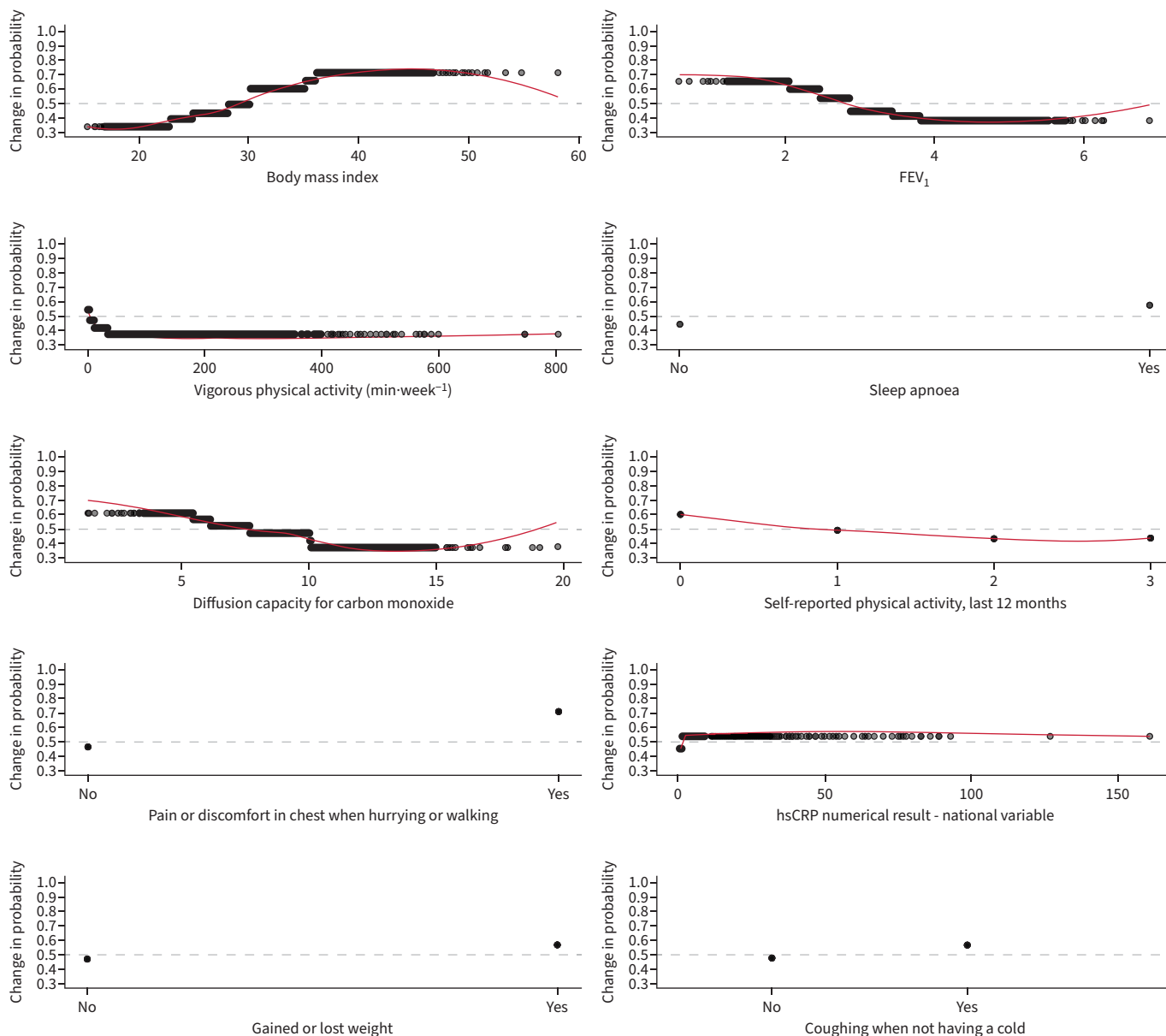


FIGURE 2 Shape of associations with breathlessness for the 10 strongest associated factors. Each dot represents a participant. The x-axis presents the values of the participants' factors. The y-axis presents the probability (0–1) of being breathless. The dashed lines indicate the probability of 0.5. A score above 0.5 on the y-axis corresponds to an association with being breathless, and a score below 0.5 corresponds to an association with not being breathless. The red lines are locally estimated scatterplot smoothing lines, which represent an average mean by factor values of the participants. Dichotomous factors are represented by two dots without lines. FEV₁: forced expiratory volume in 1 s; HsCRP: high-sensitivity C-reactive protein.

the mMRC scale in their daily life [9, 20, 21]. However, men and women were found to report similar rates of breathlessness when compared to people with similar absolute lung function (ventilatory capacity), indicating that women experience more breathlessness due to being overall smaller and having lower ventilatory capacity compared to men [9, 21]. Consistently, the present findings show that the difference in breathlessness between men and women was related to absolute lung function and other factors in the analysis.

Our findings showed an association between sleep apnoea and breathlessness which is not well-studied and should be investigated in future studies. A previous study has shown an association between sleep problems and breathlessness; however, the study did not report sleep apnoea specifically [22].

Implications

This study identifies several factors strongly related to breathlessness that are likely to be amenable to interventions and to be relevant public health targets for the reduction and prevention of breathlessness in the general population. Based on the present findings, suggested interventions include preventing and reducing smoking and exposure to other noxious exposures, unwanted weight gain and obesity, and promoting physical activity. A sedentary lifestyle was strongly and directly associated with breathlessness, and as increased physical activity decreases BMI, promoting physical activity should be a top priority in the middle-aged population. There is a need for complex interventions that target all of these factors on a population level, which could also be beneficial for the overall population health. An implication for clinical practice is that spirometry is an important investigation in primary care in patients experiencing breathlessness in order to diagnose respiratory diseases.

Strengths and limitations

The strengths of this study include the large, population-based sample and the machine learning approach to evaluate the complex interplay between a multitude of factors in relation to breathlessness. Machine learning methods can account for multiple factors and nonlinear associations, and the shape of the associations with breathlessness was evaluated. The prevalence of obesity was very similar to the mean obesity prevalence in the Organisation for Economic Cooperation and Development (OECD) countries [23], which increases the generalisability of our findings. The strengths are balanced by some limitations, as the cross-sectional study design prevents the causative evaluation of factors and the narrow age span of 50–64 years limits the generalisability to younger or older individuals. The performance of the model was acceptable (rather than excellent), but it is challenging to reach a very high classification performance in a dataset with a low prevalence of the outcomes, which explains the high sensitivity and slightly lower specificity. The anxiety and depression measurements have some limitations [24], and it is possible that these factors would have been more strongly associated with breathlessness if they had more robust measurements [1]. The mMRC is reliable, [25] strongly related to clinical outcomes [26] and commonly used in population studies, which simplifies comparison with previous findings, but the scale is limited in that it only measures the degree of exertional limitations due to breathlessness by recall and neither in relation to a standardised level of exertion nor the intensity (or severity) of the symptom *per se* [27]. Future studies should consider measuring the association between factors and standardised levels of exertion, intensity and other dimensions of breathlessness, as the factors can be different or have different strengths of association with multidimensional measurements of breathlessness compared to exertional breathlessness [28].

Conclusion

This large population-based study of men and women aged 50–64 years identified increased BMI, reduced lung function and a sedentary lifestyle as the main factors associated with breathlessness. At least two of these factors can be prevented or are amenable to public health interventions.

Provenance: Submitted article, peer reviewed.

Ethics statement: The SCAPIS multicentre study was approved by the ethics committee at Umeå University (Dnr 2010–228-31 M) and the present analysis was approved by the Swedish Ethical Review Authority (reference number 2021-00288). All participants gave their written informed consent.

Conflict of interest: M. Olsson, A.J. Björkelund, J. Sandberg, A. Blomberg, A. Malinowski, M. Börjesson, M. Sköld, P. Wollmer, K. Torén, C.J. Östgren, G. Engström and Magnus Ekström report no conflict of interest related to this work. D. Currow reports receiving consulting fees from Helsinn Pharmaceuticals, Mayne Pharma International Pty Ltd, Nous Group Pty Ltd, Icare Dust Diseases Board, Chris O'Brien Lifehouse and IHMRI, outside the submitted work; and received payment for intellectual property from Mayne Pharma International Pty Ltd.

Support statement: The main funding body of The Swedish Cardiopulmonary Bioimage Study (SCAPIS) is the Swedish Heart–Lung Foundation. The study is also funded by the Knut and Alice Wallenberg Foundation, the Swedish Research Council, VINNOVA (Sweden's innovation agency), the University of Gothenburg and Sahlgrenska University Hospital, Karolinska Institutet, Stockholm county council, Linköping University and University Hospital, Lund University and Skåne University Hospital, Umeå University and University Hospital, and Uppsala University and University Hospital. M. Ekström and M. Olsson were supported by unrestricted grants from the Swedish Research Council (ref. 2019-02081). Funding information for this article has been deposited with the Crossref Funder Registry.

References

- 1 Sandberg J, Olsson M, Ekstrom M. Underlying conditions contributing to breathlessness in the population. *Curr Opin Support Palliat Care* 2021; 15: 219–225.
- 2 Ekström M, Sundh J, Andersson A, et al. Exertional breathlessness related to medical conditions in middle-aged people: the population-based SCAPIS study of more than 25,000 men and women. *Respir Res* 2024; 25: 127.
- 3 Pedersen F, Mehlsen J, Raymond I, et al. Evaluation of dyspnoea in a sample of elderly subjects recruited from general practice. *Int J Clin Pract* 2007; 61: 1481–1491.
- 4 Lundberg SM, Erion G, Chen H, et al. From Local explanations to global understanding with explainable AI for trees. *Nat Mach Intell* 2020; 2: 56–67.
- 5 Bergstrom G, Berglund G, Blomberg A, et al. The Swedish CARDioPulmonary Biolmage Study: objectives and design. *J Intern Med* 2015; 278: 645–659.
- 6 Bergstrom G, Persson M, Adiels M, et al. Prevalence of subclinical coronary artery atherosclerosis in the general population. *Circulation* 2021; 144: 916–929.
- 7 Mahler DA, Wells CK. Evaluation of clinical methods for rating dyspnea. *Chest* 1988; 93: 580–586.
- 8 Statistics Sweden. Date last accessed: 3 April 2023. www.scb.se/en/
- 9 Ekstrom M, Sundh J, Schioler L, et al. Absolute lung size and the sex difference in breathlessness in the general population. *PLoS One* 2018; 13: e0190876.
- 10 Ekström M. Obesity is a major contributing cause of breathlessness in the population. *Respirology* 2023; 28: 303–304.
- 11 Olsson M, Currow DC, Ekstrom MP. Exploring the most important factors related to self-perceived health among older men in Sweden: a cross-sectional study using machine learning. *BMJ Open* 2022; 12: e061242.
- 12 Liu Y, Chen PC, Krause J, et al. How to read articles that use machine learning: users' guides to the medical literature. *JAMA* 2019; 322: 1806–1816.
- 13 Chen T, Guestrin C. XGBoost: A scalable tree boosting system. *arXiv* 2016; preprint [<https://doi.org/10.48550/arXiv.1603.02754>].
- 14 Guo YL, Ampon MR, Poulos LM, et al. Contribution of obesity to breathlessness in a large nationally representative sample of Australian adults. *Respirology* 2023; 28: 350–356.
- 15 Ekström MP, Blomberg A, Bergström G, et al. The association of body mass index, weight gain and central obesity with activity-related breathlessness: the Swedish Cardiopulmonary Bioimage Study. *Thorax* 2019; 74: 958.
- 16 Kochovska S, Currow D, Chang S, et al. Persisting breathlessness and activities reduced or ceased: a population study in older men. *BMJ Open Respir Res* 2022; 9: e001168.
- 17 Johnson MJ, Yorke J, Hansen-Flaschen J, et al. Towards an expert consensus to delineate a clinical syndrome of chronic breathlessness. *Eur Respir J* 2017; 49: 1602277.
- 18 Jakeways N, McKeever T, Lewis SA, et al. Relationship between FEV₁ reduction and respiratory symptoms in the general population. *Eur Respir J* 2003; 21: 658–663.
- 19 Díaz AA, Morales A, Díaz JC, et al. CT and physiologic determinants of dyspnea and exercise capacity during the six-minute walk test in mild COPD. *Respir Med* 2013; 107: 570–579.
- 20 Gronseth R, Vollmer WM, Hardie JA, et al. Predictors of dyspnoea prevalence: results from the BOLD study. *Eur Respir J* 2014; 43: 1610–1620.
- 21 Ekstrom M, Schioler L, Gronseth R, et al. Absolute values of lung function explain the sex difference in breathlessness in the general population. *Eur Respir J* 2017; 49: 1602047.
- 22 Currow DC, Chang S, Ferreira D, et al. Chronic breathlessness and sleep problems: a population-based survey. *BMJ Open* 2021; 11: e046425.
- 23 Blüher M. Obesity: global epidemiology and pathogenesis. *Nat Rev Endocrinol* 2019; 15: 288–298.
- 24 Sandberg J, Ekstrom M, Borjesson M, et al. Underlying contributing conditions to breathlessness among middle-aged individuals in the general population: a cross-sectional study. *BMJ Open Respir Res* 2020; 7: e000643.
- 25 Mahler DA, Ward J, Waterman LA, et al. Patient-reported dyspnea in COPD reliability and association with stage of disease. *Chest* 2009; 136: 1473–1479.
- 26 Nishimura K, Izumi T, Tsukino M, et al. Dyspnea is a better predictor of 5-year survival than airway obstruction in patients with COPD. *Chest* 2002; 121: 1434–1440.
- 27 Laviolette L, Laveneziana P, ERS Research Seminar Faculty. Dyspnoea: a multidimensional and multidisciplinary approach. *Eur Respir J* 2014; 43: 1750–1762.
- 28 Olsson M, Currow DC, Johnson MJ, et al. Prevalence and severity of differing dimensions of breathlessness among elderly males in the population. *ERJ Open Res* 2022; 8: 00553-2021.