bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Heterogeneity in the respiratory symptoms of patients with mild-moderate COPD

Kate M. Johnson[1], Abdollah Safari[1,4], Wan C. Tan[2], Jean Bourbeau[3], J Mark FitzGerald[4], Mohsen Sadatsafavi[1,4,5], for the *Canadian Cohort of Obstructive Lung Disease (CanCOLD) study* and the *Canadian Respiratory Research Network*

Affiliations:

1. Collaboration for Outcomes Research and Evaluation, Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, Canada
2. Centre for Heart Lung Innovation (the James Hogg Research Centre), St. Paul's Hospital, Vancouver, Canada
3. Respiratory Epidemiology and Clinical Research Unit, McGill University Health Centre, Montreal, Canada
4. Institute for Heart and Lung Health, Department of Medicine, the University of British Columbia, Vancouver, Canada
5. Centre for Clinical Epidemiology and Evaluation, Vancouver Coastal Health Institute, Vancouver, Canada

Corresponding Author: Mohsen Sadatsafavi
Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver Campus,
2405 Wesbrook Mall
Vancouver, BC Canada V6T 1Z3
Tel: 604.827.3020 | Fax: 604.875.5179
Email: msafavi@mail.ubc.ca

Running header: Heterogeneity in COPD symptoms
Word count: 2,783 (without abstract, references, tables, and figure captions)

bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

ABSTRACT

Background: The burden of symptoms varies markedly between patients with Chronic Obstructive Pulmonary Disease (COPD) and is only weakly correlated with lung function impairment. While heterogeneity in lung function decline and exacerbations have been previously studied, the extent of heterogeneity in symptoms and the factors associated with this heterogeneity are not well understood.

Methods: A sample of the general Canadian population ≥40 years with persistent airflow limitation was followed for up to 3 years. Participants reported whether they experienced chronic coughing, phlegm, wheezing, or dyspnea during visits at 18-month intervals. We used mixed-effect logistic regression models (separately for each symptom) to assess overall heterogeneity in the occurrence of symptoms between individuals, and the proportion of variation in symptom burden explained by lung function versus all other clinical characteristics of participants.

Results: 548 participants (54% male, mean age 67 years) contributed 1,086 visits in total, and 82% of patients reported at least one symptom during follow-up. There was substantial heterogeneity in the individual-specific probabilities for the occurrence of symptoms. This heterogeneity was highest for dyspnea and lowest for phlegm (interquartile range of probabilities: 0.15-0.77 and <0.01-0.53, respectively). FEV_1 explained 82% of the variation between individuals in the occurrence of phlegm, 26% for dyspnea, 3% for cough, and <0.1% for wheeze. All clinical characteristics of participants (including FEV_1) explained between 86% of heterogeneity in the occurrence of phlegm to <1% for wheeze.

Conclusion: There is marked heterogeneity in the burden of respiratory symptoms between COPD patients. The ability of lung function and other commonly measured clinical characteristics to explain this heterogeneity differs between symptoms.

bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

25 **Word count:** 262

26

27 **Keywords:** Population; Respiratory symptoms; Chronic Obstructive Pulmonary Disease;

28 Variability

bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1.0 INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a common inflammatory lung condition that affects close to 400 million people worldwide.[1] COPD is characterized by persistent airflow limitation and symptoms such as breathlessness, chronic cough, sputum production, wheezing, and chest tightness.[2] Respiratory symptoms are a major burden in many patients, and are associated with an increased frequency of exacerbations,[3] worse disease prognosis,[4–6] lower health status,[7,8] reduced quality of life,[9] and higher healthcare resource utilization.[10]

The three major components of the natural history of COPD are lung function status, patterns of exacerbations, and symptom burden.[2] The degree of symptom impairment is increasingly recognized as an important determinant of patient management strategies, and one that is only partially dependent on the severity of airflow limitation.[4,7,11,12] Indeed, modern guidelines such as the Global Initiative for Chronic Obstructive Lung Disease (GOLD) appreciate the importance of all three components in disease management decisions. The GOLD guidelines recommend evaluating symptoms separately from airflow limitation and history of exacerbations in providing therapeutic recommendations.[2]

It is increasingly recognized that COPD is a heterogeneous disease. Individuals can vary markedly in their rate of lung function decline[13] and frequency of exacerbations[14,15] over the course of their disease. For example, COPD patients in the Lung Health Study had an annual rate of change in FEV_1 that ranged from rapidly declining to modestly increasing (95% CI -83 mL/yr to +15 mL/yr).[13] Similarly, the annual rate of exacerbations observed in the MACRO clinical trial varied from 0.47 to 4.22.[15] Quantifying this variation at an individual level is critical to enabling precise risk factor and disease management.[16]

In contrast, heterogeneity in the burden of symptoms has not received the same level of attention as these other disease components. Previous studies have reported that patient

bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

symptoms tend to vary over the day, week, or season,[11,17–19] but the extent of variation between individuals in the occurrence of symptoms has been less well characterized. Understanding the extent and drivers of this heterogeneity can help improve our understanding of the natural history of COPD and ultimately help formulate disease management strategies that provide optimal therapeutic strategies for each patient.

Using data from a population-based prospective cohort, we assessed the burden of self-reported respiratory symptoms in patients with persistent airflow limitation in order to (1) characterize variation in the occurrence of symptoms between individuals, and (2) determine the proportion of between-individual variability in symptoms that can be explained by lung function versus all other observable characteristics. We hypothesized that there is high variability in the occurrence of symptoms between individuals, and that an individual's clinical and demographic characteristics explain a larger fraction of this heterogeneity than lung function alone.

2.0 METHODS

We used data from the Canadian Cohort of Obstructive Lung Disease (CanCOLD), which is a multicenter prospective longitudinal cohort study conducted across Canada.[20] Individuals ≥40 years old were recruited using random digit dialing and multi-level sampling to ensure representativeness of the general Canadian population. Participants were followed for a maximum of 3 years with in-person visits at baseline and at 18-month intervals. From the entire cohort of CanCOLD participants (N=1,532), we selected all visits in which the patient had persistent airflow limitation, defined as post-bronchodilator FEV1/FVC< lower limit of normal (LLN).[21]

bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Information was collected during each visit on the presence of cough, phlegm, wheeze, and dyspnea using separate questions for each symptom. Participants reported whether they (1) usually coughed in the absence of a cold, (2) brought up phlegm from the chest in the absence of a cold, and (3) experienced any wheezing or whistling in the chest. Dyspnea (4) was measured using the Medical Research Council (MRC) dyspnea scale,[22] which was converted to a binary variable by assuming that a score of 2-5 indicated the presence of dyspnea. Dyspnea and whether the participant experienced any symptoms were assessed in a subset of the data that included 957 visits from 502 participants because 46 participants were unable to walk and therefore did complete the MRC dyspnea test. Participants reported their demographic information, smoking status and history, number of comorbidities, and history of physician-diagnosed COPD, emphysema, or chronic bronchitis at each visit using validated questionnaires with a recall period spanning the length of time between visits.[23]

Statistical analysis

We used separate random effect logistic regression models for cough, phlegm, wheeze, dyspnea, and any symptoms to model heterogeneity. The random effect term captured the variability among individuals (heterogeneity) that was not attributable to the independent variables in the model. We initially determined the total heterogeneity in the occurrence of symptoms using an intercept-only logistic regression model for each symptom (the null model, *ie*, no independent variables). We used this model to determine the individual-specific probability of experiencing each symptom, and estimated the interquartile (25%-75%) range of probabilities to measure heterogeneity in the occurrence of symptoms.

bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

We subsequently assessed the proportion of the total heterogeneity in symptoms that could be explained by all measured characteristics of individuals. For this, we included patient age, sex, body mass index (BMI), ethnicity, number of comorbidities, smoking status, pack-years of smoking, post-bronchodilator forced expiratory volume in 1 second (FEV_1), and diagnosis status as independent variables in each model (the full models). In order to determine the variance explained by the independent variables (*ie,* participants' measured characteristics), we calculated the proportion of variance in the random effect of the full model for each symptom (with all the independent variables), compared to the variance in the random effect of the null model with no independent variables.[24] We repeated this process using a reduced model with FEV_1 as the only independent variable (as opposed to the full model) to determine the proportion of total heterogeneity explained by lung function alone.

We conducted sensitivity analyses in which percent predicted FEV_1 and GOLD grade were used in place of FEV_1 as indicators of lung function (collinearity prevented these variables from being included in the model at the same time). Seasonality was not included in the main analysis because the recall period was >1 year and therefore spanned all seasons; however, season was assessed in a sensitivity analysis to account for the possibility that patients were more likely to recall their recent symptom burden (which could be affected by the current season). All analyses were performed in SAS (version 9.4, 2016).

3.0 RESULTS

The characteristics of participants are shown in ***Table 1***. There were 1,086 visits from 548 participants in the final sample (54% male, mean age 67 years). 92% of participants had mild to

bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

121 moderate disease (grade I-II), 7% had severe disease (grade III), and 1% had very severe disease
122 (grade IV) as measured by GOLD grades.[2] 74% of participants with persistent airflow limitation
123 on spirometry had not been previously diagnosed. The average follow-up time was 18 months;
124 39% of participants underwent only one study visit, and 37% of participants were assessed at
125 three study visits. The characteristics of the subset of the data used to analyze dyspnea and any
126 symptoms were very similar (***Supplementary Material, Table 1***).

127 <<Table 1>>

128

129 ***Objective 1: Heterogeneity in the occurrence of symptoms***

130 Most participants did not report having cough, phlegm, wheeze, or dyspnea at each study visit,
131 but only 18% of participants were completely asymptomatic throughout the study period. The
132 asymptomatic participants tended to have mild airflow obstruction (mean of 89% predicted
133 FEV_1, 16% SD). The proportion of patients that reported a given symptom at least once during
134 follow-up ranged from 38% for phlegm (the least common symptom) to 56% for dyspnea (the
135 most common symptom). Symptoms were generally stable within participants: 66% of
136 participants reported the same level of cough throughout their follow-up (the least stable
137 symptom), and 75% for phlegm (the most stable symptom).

138 There was substantial variation in the individual-specific probabilities for the occurrence
139 of symptoms (***Figure 1***). The median probabilities of an individual experiencing cough, wheeze,
140 and dyspnea were 0.36, 0.34, and 0.44, respectively. In contrast, the median probability of
141 experiencing phlegm was <0.01 and it was >0.99 for any symptoms. The interquartile range of
142 probabilities was 0.13-0.68 for cough, <0.01-0.53 for phlegm, 0.10-0.72 for wheeze, 0.15-0.77

bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

for dyspnea, and 0.56->0.99 for any symptoms. Median probabilities are depicted with blue lines and interquartile ranges are depicted with grey boxes in Figure 1.

<<Figure 1 >>

Objective 2: Influence of lung function on symptom heterogeneity

The logistic regression models revealed relatively consistent associations between patient and disease characteristics and the presence of cough, phlegm, wheeze, dyspnea and any symptoms. Comparisons of the strength of associations across individual symptoms are shown in ***Figure 2***, and with any symptoms in ***Figure 3***. Lung function, sex, pack-years of smoking, BMI, and whether the participant had previously received a diagnosis of COPD were associated with most patient-reported symptoms. Ethnicity and the number of comorbidities were not associated with symptoms. Lung function was most strongly associated with the presence of any symptoms (OR per 100 mL increase in FEV_1 0.83, 95% CI 0.77-0.88), and least strongly associated with the presence of cough (OR 0.94, 95% CI 0.90-0.98). These results were similar when lung function was assessed as percent predicted FEV_1 or GOLD grade in sensitivity analyses (results not shown). Higher pack-years of smoking, BMI, and a previous diagnosis of COPD were all associated with an increased odds of reporting most symptoms. Males were more likely than females to report the presence of cough (OR 2.12, 95% CI 1.18-3.81), phlegm (OR 6.95, 95% CI 2.74-17.59), wheeze (OR 2.86, 95% CI 1.34-6.11), and any symptoms (OR 3.16, 95% CI 1.38-7.23). Summer (vs. winter) was associated with increased reporting of cough (OR 2.02 95% CI 1.12-3.62), wheeze (OR 2.80 95% CI 1.44-5.46) and any symptoms (OR 2.89, 95% CI 1.34-6.24) when it was included in sensitivity analyses.

<<Figure 2 >>

bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

<<Figure 3 >>

The proportion of between-individual variation in the occurrence symptoms that could be attributed to participants' measured characteristics (all independent variables in the full models) was 21%, 86%, <1%, 44%, and 92%, for cough, phlegm, wheeze, dyspnea, and any symptoms, respectively (***Table 2***). The proportion of variation explained by FEV_1 alone ranged from 0% (for wheeze) to 82% (for phlegm, ***Table 2***).

<<Table 2>>

4.0 DISCUSSION

We have characterized heterogeneity in the occurrence of respiratory symptoms between patients with persistent airflow limitation and assessed the extent to which commonly measured patient and disease characteristics explained the observed heterogeneity in symptoms. Respiratory symptoms were very common in this sample of the general population; four out of every five participants reported experiencing symptoms despite over 90% of patients having mild to moderate COPD, and only 26% of them having been diagnosed with COPD. Dyspnea was the most common symptom, followed by cough and wheeze. Individual-specific probabilities for the occurrence of symptoms were highly variable between individuals and for different symptoms. The interquartile range of probabilities was the largest for dyspnea and wheeze, indicating greater variability between individuals in the presence of these symptoms than for cough and phlegm. For phlegm, the majority of individuals had a probability of experiencing phlegm near 0 or 1 (visible in Figure 1 as a higher density of points at the edges of the plot). In contrast, the

bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

189 individual-specific probabilities for cough, wheeze, and dyspnea were more evenly spread across
190 the range of possible values. This indicates that phlegm is more stable in nature, and that
191 individuals who do not currently have phlegm are unlikely to report it in the future. Indeed, a
192 pan-European study reported that daily and weekly variability in dyspnea, wheeze, and cough
193 were higher than that for phlegm.[25] Our findings extend these observations on symptom
194 variability within individuals to variability between individuals in the occurrence of symptoms.
195 As a result, tools for assessing COPD severity that involve the measurement of symptoms (such
196 as the GOLD ABCD assessment tool)[2], are likely to be more or less variable over time,
197 depending on the symptom measured.

198 The proportion of heterogeneity explained by the measured characteristics of participants
199 differed substantially between symptoms. Most heterogeneity in the occurrence of phlegm and
200 any symptoms, and approximately half the heterogeneity in dyspnea, was explained by the
201 demographic and clinical characteristics of participants included in the models. In contrast, these
202 characteristics explained very little heterogeneity in cough and wheeze, indicating that other
203 characteristics not included in our models are more important drivers of these symptoms. Indeed,
204 age, sex, BMI, smoking history, and lung function (cough only) were weakly correlated with
205 cough frequency[26] and the presence of wheezing[27] in previous studies. Instead, cough frequency
206 was driven by current smoking intensity and percentage of sputum neutrophils,[26] and the
207 presence of wheezing was associated with frequent exacerbations and increased dyspnea.[27]

208 Although lung function has traditionally been regarded as the primary driver of
209 respiratory symptoms,[28] we found that FEV_1 explained the majority of between-individual
210 variation in only the occurrence of phlegm, although it explained a substantial minority of
211 variation in dyspnea and any symptoms. This finding is in line with the observation of high

bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

symptom variability within levels of disease severity,[29] and high short-term variability in symptoms that is not due to changes in lung function.[4,12] Our results extend these previous studies by examining the role of FEV_1 in each symptom individually. They suggest that lung function is the primary driver of the occurrence of phlegm, an important but not dominant driver of the occurrence of dyspnea and any symptoms, but it explains very little between-individual variation in the occurrence of cough and wheeze.

In addition to analyzing heterogeneity in symptoms, we documented associations between symptoms and age, sex, BMI, smoking pack-years, FEV_1, and a previous diagnosis of COPD. In particular, we observed substantial sex-based differences in the reporting of all symptoms apart from dyspnea. Controlled for disease severity, smoking history, and other variables, male patients were over three times more likely to be symptomatic, and seven times more likely to report experiencing phlegm. Whether this is a biological phenomenon, or due to gender-related differences in the experience of symptoms,[30] remains to be further assessed.

Unique features of this study are its population-based sample and our assessment of the different sources of variation in symptoms. The associations determined from conventional regressions describe the relation between patient and disease characteristics and the presence of a symptom for an average participant. Our use of a random effect term in our models enabled us to extend these results by describing the extent to which these population-level associations apply to a given individual. We found that variation between individuals in the presence of phlegm and any symptoms was very well described by these population-level associations, but this was not the case for wheeze and cough (and somewhat the case for dyspnea). The assessment of variation at an individual-level is critical to fully characterizing heterogeneity in the natural history of

bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

COPD, and ultimately to enabling effective use of symptoms in risk prediction tools and case finding algorithms for COPD.

This study has several strengths. First, CanCOLD is a large, nationally representative sample of Canadians with COPD in the general population. The study employed standardized spirometry, validated questionnaires, and a long follow-up time. Our sample consisted primarily of patients with mild-moderate COPD, a population that is often underrepresented in large cohort studies. The population-based nature of the study makes it a better source for studying disease heterogeneity than clinical cohorts. However, this study also has several limitations. Patients reported their respiratory symptoms with a recall period that spanned the length of time between study visits, which could reach a maximum of three years. The long duration of the recall period is likely to have resulted in inaccuracies in symptom reporting. In addition, we only assessed the presence of symptoms, not their intensity. A more granular measurement of patient symptoms could provide a more nuanced assessment of symptom variability. Finally, we could not investigate the impact of factors that were not included in our model, in particular the use of treatment to control symptoms. Although this would likely have reduced the proportion of variation in symptoms that we attributed to unmeasured characteristics, the long recall period is less influenced by short-term variation in symptoms due to treatment.

5.0 CONCLUSION

We assessed a sample of the general population with mostly mild-moderate COPD and found substantial variation in the occurrence of respiratory symptoms between individuals. Lung function explained the majority of between-individual variation in only the occurrence of

bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

257 phlegm, and a much smaller proportion of variation in dyspnea, cough, and wheeze. For cough
258 and wheeze in particular, commonly measured patient and disease characteristics explained very
259 little heterogeneity in the occurrence of these symptoms. Overall, the observed differences in
260 symptom variation may reflect the divergent etiology of symptoms associated with COPD, and
261 suggests that defining endotypes that are predictive of a high symptom burden is a key area of
262 future research.

bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

ABBREVIATIONS

COPD: Chronic Obstructive Pulmonary Disease

CanCOLD: Canadian Cohort of Obstructive Lung Disease Study

FEV_1: Forced Expiratory Volume in 1 second

FVC: Forced Vital Capacity

GOLD: Global Initiative for chronic Obstructive Lung Disease

LLN: Lower Limit of Normal

MRC: Medical Research Council dyspnea scale

OR: Odds Ratio

SD: Standard Deviation

DECLARATIONS

Ethics approval and informed consent: Ethics approval for CanCOLD was obtained from the relevant institutional review board at each study site. Written informed consent was obtained from all participants prior to study entry.

Consent for publication: Not applicable.

Availability of data and materials: The data analyzed in the current study are not publicly available but may be made available from the CanCOLD Research Group upon reasonable request.

Competing interests: The authors declare that they have no competing interests.

Funding: The current study was funded by a Canadian Lung Association Breathing as One Studentship Award and the Canadian Institutes of Health Research (application number 142238). The Canadian Cohort Obstructive Lung Disease (CanCOLD) study is currently funded by the

bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Canadian Respiratory Research Network (CRRN); industry partners: Astra Zeneca Canada Ltd; Boehringer Ingelheim Canada Ltd; GlaxoSmithKline Canada Ltd; and Novartis. Researchers at RI-MUHC Montreal and Icapture Centre Vancouver lead the project. Previous funding partners are the CIHR (CIHR/Rx&D Collaborative Research Program Operating Grants 93326); the Respiratory Health Network of the Fonds de la recherche en santé du Québec (FRSQ); industry partners: Almirall; Merck Nycomed; Pfizer Canada Ltd; and Theratechnologies. The funders had no role in study design, data collection and analysis, or preparation of the manuscript.

Author Contributions: WT and JB are co-Principal Investigators of the CanCOLD study. MS and KJ formulated the current study idea. KJ performed all data analyses and wrote the first draft of the manuscript. AS provided guidance on the statistical analysis. All authors contributed to interpretation of findings, critically commented on the manuscript and approved the final version. MS is the guarantor of the manuscript.

Acknowledgements: The authors thank the men and women who participated in the study and individuals in the CanCOLD Collaborative Research Group.

Members of the CanCOLD Collaborative Research Group are as follows. Executive Committee: Jean Bourbeau (McGill University, Montreal, Canada); Wan C. Tan, J. Mark FitzGerald; Don Sin (UBC, Vancouver, Canada); Darcy Marciniuk (University of Saskatoon, Saskatoon, Canada); Dennis E. O'Donnell (Queen's University, Kingston, Canada); Paul Hernandez (Dalhousie University, Halifax, Canada); Kenneth R. Chapman (University of Toronto, Toronto, Canada); Robert Cowie (University of Calgary, Calgary, Canada); Shawn Aaron (University of Ottawa, Ottawa, Canada); F. Maltais (University of Laval, Quebec City, Canada). International Advisory Board: Jonathon Samet (Keck School of Medicine of USC, Los Angeles, CA); Milo Puhan (John Hopkins School of Public Health, Baltimore, MD); Qutayba Hamid (McGill

bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

University, Montreal, Canada); James C. Hogg (UBC James Hogg Research Center, Vancouver, Canada). Operations Center: Jean Bourbeau (PI), Carole Jabet, Palmina Mancino, (McGill University, Montreal, Canada); Wan C. Tan (co-PI), Don Sin, Sheena Tam, Jeremy Road, Joe Comeau, Adrian Png, Harvey Coxson, Jonathon Leipsic, Cameron Hague (University of British Columbia James Hogg Research Center, Vancouver, Canada). Economic Core: Mohsen Sadatsafavi (University of British Columbia, Vancouver, Canada). Public Health Core: Teresa To, Andrea Gershon (University of Toronto, Toronto, Canada). Data Management and Quality Control: Wan C. Tan, Harvey Coxson (UBC, Vancouver, Canada); Jean Bourbeau, Pei Zhi Li, Zhi Song, Yvan Fortier, Andrea Benedetti, Dennis Jensen (McGill University, Montreal, Canada). Field Centers: Wan C. Tan (Vancouver PI), Christine Lo, Sarah Cheng, Elena Un, Cindy Fung, Nancy Haynes, Junior Chuang, Licong Li, Selva Bayat, Amanda Wong, Zoe Alavi, Catherine Peng, Bin Zhao, Nathalie Scott-Hsiung, Tasha Nadirshaw (UBC James Hogg Research Center, Vancouver, Canada); Jean Bourbeau (Montreal PI), Palmina Mancino, David Latreille, Jacinthe Baril, Laura Labonté (McGill University, Montreal, Canada); Kenneth Chapman (Toronto PI), Patricia McClean, Nadeen Audisho (University of Toronto, Toronto, Canada); R. Cowie and B. Walter (Calgary PI), Ann Cowie, Curtis Dumonceaux, Lisette Machado (University of Calgary, Calgary, Canada); Paul Hernandez (Halifax PI), Scott Fulton, Kristen Osterling (Dalhousie University, Halifax, Canada); Shawn Aaron (Ottawa PI), Kathy Vandemheen, Gay Pratt, Amanda Bergeron (University of Ottawa, Ottawa, Canada); Denis O'Donnell (Kingston PI), Matthew McNeil, Kate Whelan (Queen's University, Kingston, Canada); François Maltais (Quebec PI), Cynthia Brouillard (Université Laval, Quebec City, Canada); Darcy Marciniuk (Saskatoon PI), Ron Clemens, Janet Baran (University of Saskatoon, Saskatoon, Canada).

bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

REFERENCES

1. Adeloye D, Chua S, Lee C, et al. Global and regional estimates of COPD prevalence: Systematic review and meta-analysis. *J Glob Health*. 2015;5(2):020415. doi:10.7189/jogh.05-020415

2. From the Global Strategy for the Diagnosis, Management and Prevention of COPD. Global Initiative for Chronic Obstructive Lung Disease (GOLD). 2017. http://goldcopd.org. Accessed May 1, 2017.

3. Roche N, Small M, Broomfield S, Higgins V, Pollard R. Real world COPD: association of morning symptoms with clinical and patient reported outcomes. *COPD*. 2013;10(6):679-686. doi:10.3109/15412555.2013.844784

4. Miravitlles M, Ribera A. Understanding the impact of symptoms on the burden of COPD. *Respir Res*. 2017;18:67. doi:10.1186/s12931-017-0548-3

5. Bridevaux P-O, Gerbase MW, Probst-Hensch NM, Schindler C, Gaspoz J-M, Rochat T. Long-term decline in lung function, utilisation of care and quality of life in modified GOLD stage 1 COPD. *Thorax*. 2008;63(9):768-774. doi:10.1136/thx.2007.093724

6. Rennard SI, Drummond MB. Early chronic obstructive pulmonary disease: definition, assessment, and prevention. *Lancet*. 2015;385(9979):1778-1788. doi:10.1016/S0140-6736(15)60647-X

7. Miravitlles M, Worth H, Soler Cataluña JJ, et al. Observational study to characterise 24-hour COPD symptoms and their relationship with patient-reported outcomes: results from the ASSESS study. *Respir Res*. 2014;15:122. doi:10.1186/s12931-014-0122-1

8. Ioanna T, Kocks J, Tzanakis N, Siafakas N, van der Molen T. Factors that influence disease-specific quality of life or health status in patients with COPD: a systematic review and meta-analysis of Pearson correlations. *Prim Care Respir J*. 2011;20(3):257-268. doi:10.4104/pcrj.2011.00029

9. Monteagudo M, Rodríguez-Blanco T, Llagostera M, et al. Factors associated with changes in quality of life of COPD patients: a prospective study in primary care. *Respir Med*. 2013;107(10):1589-1597. doi:10.1016/j.rmed.2013.05.009

10. Price D, Small M, Milligan G, Higgins V, Gil EG, Estruch J. Impact of night-time symptoms in COPD: a real-world study in five European countries. *Int J Chron Obstruct Pulmon Dis*. 2013;8:595-603. doi:10.2147/COPD.S48570

11. Lopez-Campos JL, Calero C, Quintana-Gallego E. Symptom variability in COPD: a narrative review. *Int J Chron Obstruct Pulmon Dis*. 2013;8:231-238. doi:10.2147/COPD.S42866

bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

12. Miravitlles M, Menezes A, López Varela MV, et al. Prevalence and impact of respiratory symptoms in a population of patients with COPD in Latin America: The LASSYC observational study. *Respir Med*. 2018;134:62-69. doi:10.1016/j.rmed.2017.11.018

13. Zafari Z, Sin DD, Postma DS, et al. Individualized prediction of lung-function decline in chronic obstructive pulmonary disease. *CMAJ*. 2016;188(14):1004-1011. doi:10.1503/cmaj.151483

14. Hurst JR, Vestbo J, Anzueto A, et al. Susceptibility to Exacerbation in Chronic Obstructive Pulmonary Disease. *N Engl J Med*. 2010;363(12):1128-1138. doi:10.1056/NEJMoa0909883

15. Sadatsafavi M, Sin DD, Zafari Z, et al. The Association Between Rate and Severity of Exacerbations in Chronic Obstructive Pulmonary Disease: An Application of a Joint Frailty-Logistic Model. *Am J Epidemiol*. 2016;184(9):681-689. doi:10.1093/aje/kww085

16. Jameson JL, Longo DL. Precision Medicine — Personalized, Problematic, and Promising. *N Engl J Med*. 2015;372(23):2229-2234. doi:10.1056/NEJMsb1503104

17. Partridge MR, Karlsson N, Small IR. Patient insight into the impact of chronic obstructive pulmonary disease in the morning: an internet survey. *Curr Med Res Opin*. 2009;25(8):2043-2048. doi:10.1185/03007990903103006

18. Espinosa de los Monteros MJ, Peña C, Soto Hurtado EJ, Jareño J, Miravitlles M. Variability of respiratory symptoms in severe COPD. *Arch Bronconeumol*. 2012;48(1):3-7. doi:10.1016/j.arbres.2011.07.004

19. Donaldson GC, Goldring JJ, Wedzicha JA. Influence of season on exacerbation characteristics in patients with COPD. *Chest*. 2012;141(1):94-100. doi:10.1378/chest.11-0281

20. Bourbeau J, Tan WC, Benedetti A, et al. Canadian Cohort Obstructive Lung Disease (CanCOLD): Fulfilling the need for longitudinal observational studies in COPD. *COPD*. 2014;11(2):125-132. doi:10.3109/15412555.2012.665520

21. 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(1):179-187. doi:10.1164/ajrccm.159.1.9712108

22. Fletcher CM, Elmes PC, Fairbairn AS, Wood CH. The significance of respiratory symptoms and the diagnosis of chronic bronchitis in a working population. *Br Med J*. 1959;2(5147):257-266.

23. Buist AS, McBurnie MA, Vollmer WM, et al. International variation in the prevalence of COPD (the BOLD Study): a population-based prevalence study. *Lancet*. 2007;370(9589):741-750. doi:10.1016/S0140-6736(07)61377-4

bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

24. Ten Have TR, Localio AR. Empirical Bayes estimation of random effects parameters in mixed effects logistic regression models. *Biometrics*. 1999;55(4):1022-1029.

25. Kessler R, Partridge MR, Miravitlles M, et al. Symptom variability in patients with severe COPD: a pan-European cross-sectional study. *Eur Respir J*. 2011;37(2):264-272. doi:10.1183/09031936.00051110

26. Sumner H, Woodcock A, Kolsum U, et al. Predictors of objective cough frequency in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2013;187(9):943-949. doi:10.1164/rccm.201211-2000OC

27. Huang W-C, Tsai Y-H, Wei Y-F, et al. Wheezing, a significant clinical phenotype of COPD: experience from the Taiwan Obstructive Lung Disease Study. *Int J Chron Obstruct Pulmon Dis*. 2015;10:2121-2126. doi:10.2147/COPD.S92062

28. Fletcher C, Peto R. The natural history of chronic airflow obstruction. *Br Med J*. 1977;1(6077):1645-1648. doi:10.1136/bmj.1.6077.1645

29. Agusti A, Calverley PMA, Celli B, et al. Characterisation of COPD heterogeneity in the ECLIPSE cohort. *Respir Res*. 2010;11:122. doi:10.1186/1465-9921-11-122

30. Barsky AJ, Peekna HM, Borus JF. Somatic Symptom Reporting in Women and Men. *J Gen Intern Med*. 2001;16(4):266-275. doi:10.1046/j.1525-1497.2001.00229.x

bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 1 Characteristics of study participants at study visits. Means (and standard deviations) are reported unless otherwise indicated

	Visit 1 (n=461)	**Visit 2** (n=309)	**Visit 3** (n=316)
Age	65.5 (10.3)	67.2 (10.3)	68.2 (9.5)
Male (vs. female)	54.4%	53.4%	52.5%
BMI	27.4 (5.7)	27.6 (6.0)	27.4 (5.9)
Caucasian (vs. non-Caucasian)	97.4%	97.4%	98.1%
Comorbidities			
0 comorbidities	59.2%	86.1%	86.1%
1 comorbidity	32.5%	11.7%	13.0%
≥2 comorbidities	8.2%	2.3%	0.9%
Smoking between visits (vs. no)	71.8%	27.2%	25.0%
Lifetime pack-years smoked	25.5 (25.9)	25.2 (25.4)	25.7 (25.5)
FEV_1 (L)	2.2 (0.8)	2.1 (0.8)	2.1 (0.7)
% Predicted FEV_1	76.3 (18.3)	76.0 (19.1)	76.4 (18.9)
Diagnosed COPD (vs. undiagnosed)	22.1%	26.5%	30.4%
Symptoms (present vs. absent)			
Cough	43.6%	44.7%	38.6%
Phlegm	31.2%	30.7%	26.9%
Wheeze	46.0%	42.1%	40.5%
Dyspnea*	50.0%	44.9%	42.6%
Any symptoms*	78.9%	74.5%	70.9%

* Determined for the subset of participants in which dyspnea was measured (N=502).

BMI: Body Mass Index, FEV_1: Forced Expiratory Volume in 1 second

bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 2 Percentage of between-individual variation in symptoms explained by individual's lung function and all measured characteristics combined

	Cough	**Phlegm**	**Wheeze**	**Dyspnea**	**Any Symptoms**
FEV_1	3%	82%	0%	26%	19%
All measured characteristics*	21%	86%	<1%	44%	92%

* Visit, Age, Sex, Caucasian, BMI, Comorbidities, Smoking status, Pack-years of smoking, Diagnosis status, FEV_1

BMI: Body Mass Index, FEV_1: Forced Expiratory Volume in 1 second

bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

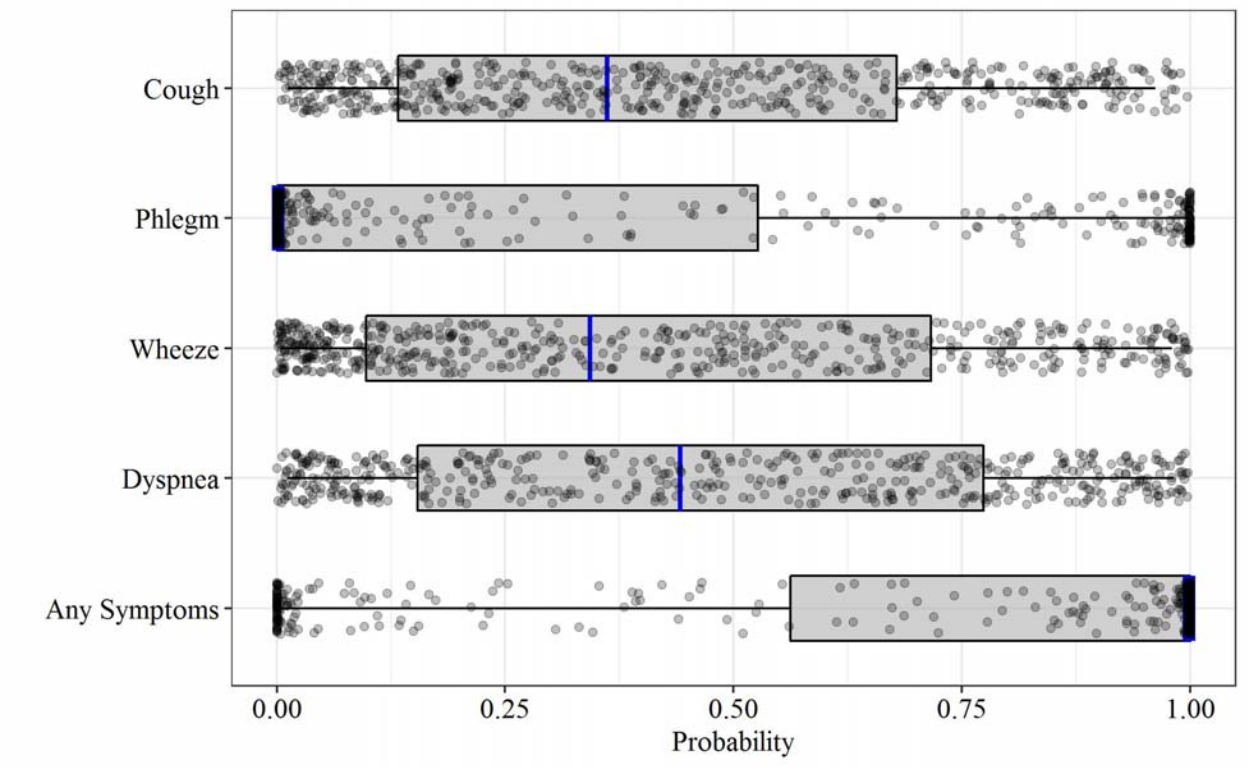


Figure 1 The distribution of individual-specific probabilities* of the occurrence of symptoms. The box spans the lower and upper quartiles (25%-75%) of individuals around the median (blue line).

*Individual random effects are drawn from a normal distribution with a mean of 0 and standard deviation of the fitted random effects. The statistics shown by the boxes were determined from 1000 repetitions for each individual, and the points show the results of one repetition.

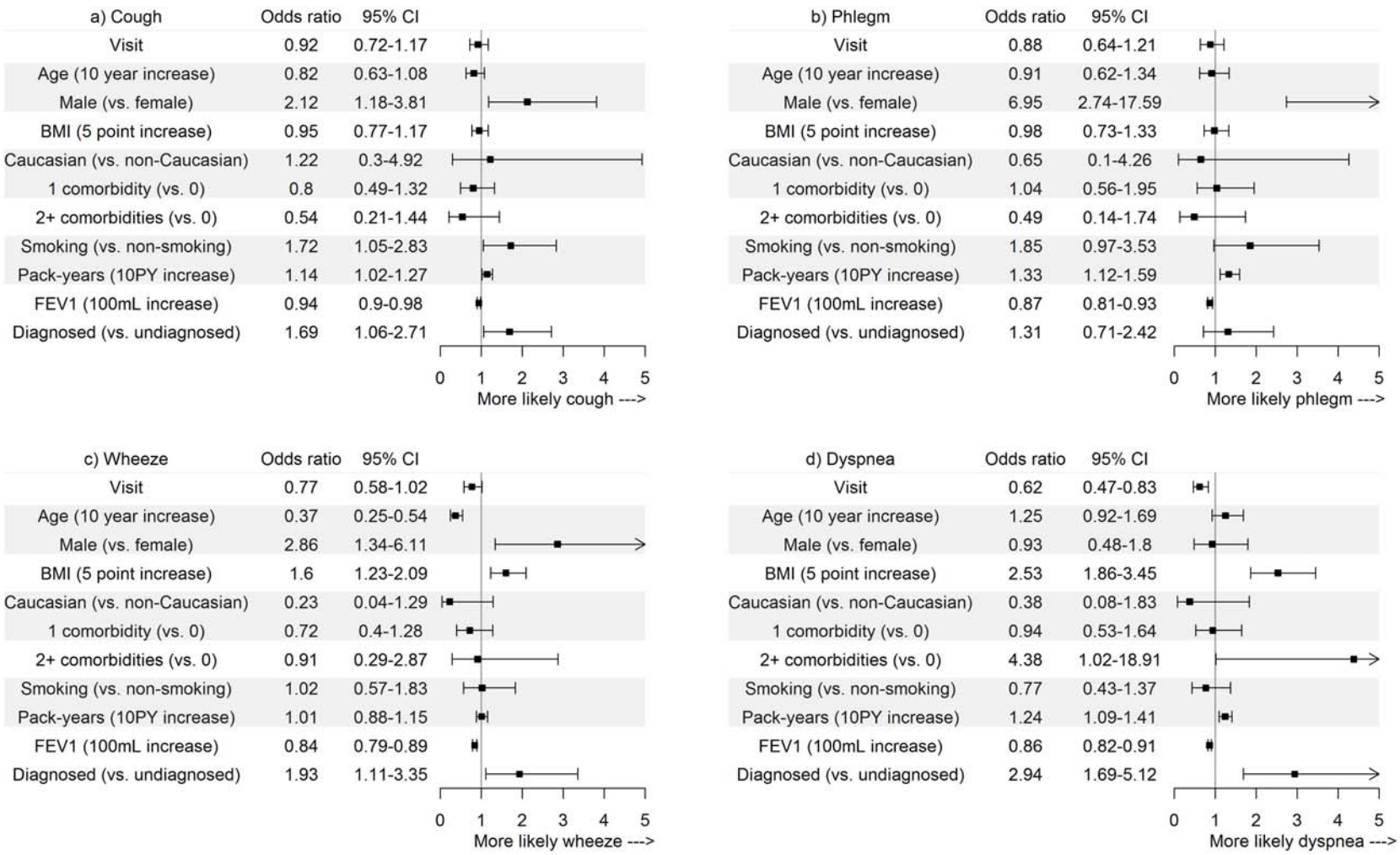


Figure 2 Odds ratios for the associations between independent variables and the presence of a) cough, b) phlegm, c) wheeze, and d) dyspnea.

bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/401695; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

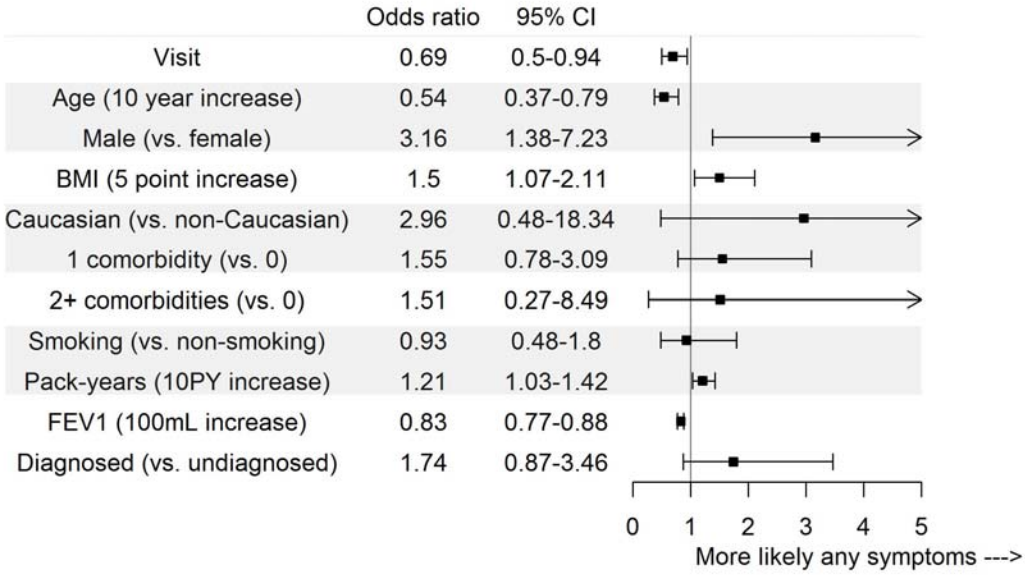


Figure 3 Odds ratios for the associations between independent variables and the presence of any symptoms.