

# Association of daily lung condition in COPD patients with wearable speech and physiological data

Sejal Bhalla<sup>1,\*</sup>, Deshang Kong<sup>1</sup>, Salaar Liaqat<sup>1</sup>, Daniyal Liaqat<sup>1</sup>, Robert Wu<sup>2</sup>, Andrea Gershon<sup>3</sup>, Eyal de Lara<sup>1</sup>, and Alex Mariakakis<sup>1</sup>

<sup>1</sup>University of Toronto, Computer Science, Toronto, ON, Canada

<sup>2</sup>University Health Network, Toronto, ON, Canada

<sup>3</sup>Sunnybrook Health Sciences Centre, Toronto, ON, Canada

\*sejal@cs.toronto.edu

## ABSTRACT

Chronic obstructive pulmonary disease (COPD) is a prevalent condition that imposes significant burden on patients and healthcare systems, with COPD exacerbations being a leading cause of hospitalizations and mortality worldwide. Continuous monitoring of lung function is crucial for effective management, yet traditional methods are often cumbersome and require clinic visits. Wearable technology offers a non-invasive alternative, enabling the monitoring of physiological parameters in real-world settings. In this study, we analyzed free-living speech data collected via smartwatches from 18 COPD patients over an average period of  $198.9 \pm 122.6$  days. Utilizing linear mixed-effects models, we discovered that phonation features are negatively associated with adverse COPD outcomes, while prosodic speech features show a positive correlation with increased exacerbation risk. Further analysis revealed that these associations are significantly moderated by physiological covariates such as heart rate variability and physical activity levels. These findings highlight the complex interplay between respiratory function, autonomic regulation, and vocal production, suggesting that the integration of speech analysis with physiological monitoring through wearables can lead to the development of composite digital biomarkers of impaired lung function.

## 1 Introduction

Chronic obstructive pulmonary disease (COPD) is the third-leading cause of death worldwide, causing 3.3 million deaths in 2019<sup>1</sup>. It is characterized by progressive airflow obstruction that results in respiratory symptoms like increased coughing and sputum production. COPD has both short- and long-term impacts on health. In the short term, symptoms can flare to the point of exacerbations, during which patients can experience decreased physical reserve and impaired quality of life<sup>2-4</sup>. In the long term, the underlying lung inflammation and subsequent exacerbations lead to multiple medical and psychological comorbidities that increase disease burden and the likelihood of hospitalization, ultimately contributing to increased mortality and an ever-growing healthcare cost<sup>5,6</sup>.

The chronic nature of COPD necessitates regular supervision to enable timely interventions that could reduce exacerbations, highlighting the need for convenient methods of continuous symptom monitoring. Although spirometry and symptom diaries have been studied to this end<sup>7-10</sup>, they only provide brief episodic glimpses into patients' health. COPD symptoms can worsen at any moment, so remote monitoring of physiological and behavioral data extracted from wearables has emerged as a promising approach to comprehensively observe patients' experiences in free-living conditions.

Beyond data related to physical activity and heart rate<sup>11-14</sup>, there has been growing interest in acoustic biomarkers of COPD because of their direct correlation with airflow limitation<sup>15-18</sup>. Speech production is a complex motor task that involves the precise coordination of orofacial, laryngeal, pharyngeal, and respiratory muscles. As a person speaks, airflow from the lungs gets modulated by the vocal folds in the larynx and is further shaped by the articulators within the vocal tract to produce sound. Given the critical role of respiratory support in this process, any impairment, such as that seen in COPD patients, can lead to notable alterations in vocal characteristics.

Prior studies have examined the feasibility of analyzing speech recordings to infer COPD outcomes such as respiratory symptoms<sup>16</sup> and lung function metrics<sup>19,20</sup>. These efforts have often involved prompted speech such as sustained vowels<sup>21,22</sup> or scripted passages of text<sup>19,22,23</sup>. Since individuals may modulate their voice when instructed to perform specific tasks, prompted speech may not capture the full range of articulatory and respiratory behaviors relevant to disease severity. These studies have also been performed in stable indoor environments, which minimizes confounding variables at the cost of ignoring the complex acoustic and physiological conditions encountered during daily life. Passive continuous monitoring throughout the day entails ambient noise, conversational variability, and varying speaking contexts, all of which may obscure subtle vocal

changes. To this end, Sedaghat et al.<sup>24</sup> proposed an end-to-end machine learning approach for inferring daily symptom severity from passively collected speech, and our recent work on the PulmoListener system<sup>25</sup> improved upon those results using deep learning. However, both of these works overlooked the influence of other factors that could have been related to symptom worsening and voice modulation, namely physical activity and stress.

The study presented in this paper addresses this shortcoming. After extracting interpretable vocal parameters from free-living speech captured using off-the-shelf smartwatches, we examine how those parameters correlate with daily self-reported symptoms and occurrences of exacerbations. Furthermore, we extend our investigation to include additional variables that may be associated with exacerbations, including step count and heart rate variability. To our knowledge, this is the first study to statistically analyze the relationship between natural speech patterns, physical activity, and heart rate variability in predicting COPD outcomes. By examining these factors concurrently, we aim to provide comprehensive insights into the multifaceted nature of COPD and contribute to the development of more effective monitoring and management strategies.

| <b>Participant Characteristics (N = 18)</b>               |                  |
|---|------------------|
| Male, n (%)   | 12 (66.67%)      |
| Age in years, mean $\pm$ SD                               | 67.3 $\pm$ 8.8   |
| Home oxygen use, n (%)                                    | 7 (38.9%)        |
| Hospital admissions in last 12 months, median (IQR)       | 2 (0–10)         |
| Days enrolled in the study, median (IQR)                  | 187.5 (24–408)   |
| <b>Comorbidities, n (%)</b>                               |                  |
| Connective tissue disease                                 | 2 (11.1%)        |
| Congestive heart failure                                  | 2 (11.1%)        |
| Diabetes  | 1 (5.6%)         |
| Mild liver disease  | 1 (5.6%)         |
| Peripheral vascular disease                               | 1 (5.6%)         |
| Ulcer disease   | 1 (5.6%)         |
| <b>Smoking History, n (%)</b>                             |                  |
| Current   | 4 (22.2%)        |
| Never   | 2 (11.1%)        |
| Ex-smoker   | 12 (66.7%)       |
| <b>COPD Outcomes</b>                                      |                  |
| Daily COPD symptom score <sup>26,27</sup> , mean $\pm$ SD | 1.3 $\pm$ 2.9    |
| Experienced at least one exacerbation, n (%)              | 6 (33.3%)        |
| Exacerbation days, mean $\pm$ SD                          | 12.8 $\pm$ 36.3  |
| <b>Wearable Data</b>                                      |                  |
| Days with speech, mean $\pm$ SD                           | 45.1 $\pm$ 50.8  |
| Daily speech duration in minutes, mean $\pm$ SD           | 122.6 $\pm$ 97.4 |
| Days with HR data, mean $\pm$ SD                          | 85.5 $\pm$ 66.9  |
| HR in beats per minute, mean $\pm$ SD                     | 86.6 $\pm$ 15.9  |
| Days with step count data, mean $\pm$ SD                  | 136.7 $\pm$ 86.4 |
| Daily step count, mean $\pm$ SD                           | 1609 $\pm$ 2069  |

**Table 1.** Participant characteristics and vital signs of COPD patients (IQR: interquartile range, SD: standard deviation)

## 2 Results

### 2.1 Participant Characteristics

The characteristics of our patient cohort are summarized in [Table 1](#). The analysis was performed with 18 patients diagnosed with COPD, 12 (66.7%) of whom were male. The participants were between 55 and 93 years old (average:  $67.3 \pm 8.8$ ). The majority had a smoking history, with 12 (66.7%) being ex-smokers and 4 (22.2%) being active smokers. Comorbidities were also prevalent among the participants, with 44.5% reporting at least one comorbidity. The most commonly reported COPD symptoms during the study were worsening breathlessness (15 patients, 83.3%), worsening cough (13 patients, 72.2%), and cold (12 patients, 66.7%). According to the London COPD Cohort Daily Symptom Questionnaire<sup>26,27</sup>, 6 patients (33.3%) experienced an exacerbation at least once during the study by recording a score exceeding 6 on two or more consecutive days.

### 2.2 Univariate associations between speech and COPD outcomes

| Category                 | Feature             | Symptom Score<br>$\beta$ -coefficient [95% CI] | Exacerbation<br>Odds Ratio [95% CI] |
|--------------------------|---------------------|--|-------------------------------------|
| Phonation Features       | HNR                 | -0.03 [-0.24, 0.17]                            | 1.35 [0.93, 1.98]                   |
|                          | jitter              | -0.20 [-0.44, 0.03]                            | 0.44 [0.25, 0.77]**                 |
|                          | absolute jitter     | -0.18 [-0.34, -0.01]*                          | 0.59 [0.42, 0.83]**                 |
|                          | ddp jitter          | -0.19 [-0.41, 0.03]                            | 0.44 [0.23, 0.83]*                  |
|                          | ppq5 jitter         | -0.16 [-0.37, 0.05]                            | 0.43 [0.24, 0.78]**                 |
|                          | rap jitter          | -0.19 [-0.41, 0.03]                            | 0.44 [0.23, 0.83]*                  |
|                          | shimmer             | -0.21 [-0.40, -0.01]*                          | 0.54 [0.37, 0.78]**                 |
|                          | apq3 shimmer        | -0.19 [-0.38, -0.00]*                          | 0.56 [0.39, 0.81]**                 |
|                          | apq5 shimmer        | -0.09 [-0.28, 0.10]                            | 0.63 [0.46, 0.87]**                 |
|                          | apq11 shimmer       | -0.07 [-0.26, 0.12]                            | 0.59 [0.41, 0.87]**                 |
|                          | dda shimmer         | -0.19 [-0.38, -0.00]*                          | 0.56 [0.39, 0.81]**                 |
| Prosodic Features        | mean F0             | 0.10 [-0.11, 0.30]                             | 1.72 [1.15, 2.57]**                 |
|                          | std F0              | 0.11 [-0.10, 0.32]                             | 1.62 [1.04, 2.53]*                  |
|                          | mean F1             | -0.05 [-0.31, 0.20]                            | 1.21 [0.67, 2.18]                   |
|                          | mean F2             | -0.24 [-0.48, -0.01]*                          | 0.61 [0.35, 1.06]                   |
|                          | mean F3             | -0.22 [-0.46, 0.02]                            | 0.77 [0.50, 1.18]                   |
|                          | mean F4             | -0.15 [-0.36, 0.06]                            | 0.77 [0.50, 1.18]                   |
|                          | mean formant        | 0.04 [-0.43, 0.51]                             | 0.76 [0.45, 1.29]                   |
|                          | delta_f             | -0.24 [-0.49, 0.02]                            | 0.75 [0.47, 1.21]                   |
|                          | fdisp               | -0.19 [-0.41, 0.03]                            | 0.66 [0.40, 1.09]                   |
|                          | fitch_vtl           | 0.04 [-0.22, 0.30]                             | 0.82 [0.47, 1.41]                   |
|                          | mff                 | -0.12 [-0.39, 0.15]                            | 0.94 [0.56, 1.61]                   |
| pF                       | -0.20 [-0.67, 0.26] | 0.98 [0.66, 1.46]                              |                                     |
| vtl_delta_f              | 0.23 [-0.02, 0.48]  | 1.32 [0.84, 2.07]                              |                                     |
| Syllabic Nuclei Features | pause rate          | -0.07 [-0.24, 0.11]                            | 1.02 [0.67, 1.54]                   |
|                          | phonation time      | 0.08 [-0.09, 0.26]                             | 0.90 [0.65, 1.24]                   |
|                          | speech rate         | 0.02 [-0.23, 0.27]                             | 0.93 [0.61, 1.42]                   |

**Table 2.** The results of the univariate linear mixed-effects models that were generated to estimate relationships between COPD outcomes and free-living speech characteristics. Significant model results are indicated as follows: \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ . Refer to [Table 6](#) for definitions of feature names and acronyms.

| (a) Symptom Score                      |  | (b) Exacerbations                      |                                      |
|--|--|--|--------------------------------------|
| Explanatory Variables                  | Symptom Score<br>$\beta$ -coefficient [95% CI] | Explanatory Variables                  | Exacerbations<br>Odds Ratio [95% CI] |
| PA(steps)                              | -0.05 [-0.28, 0.18]                            | PA(steps)                              | 1.03 [0.40, 2.68]                    |
| HRV(mean NN)                           | 0.10 [-0.16, 0.36]                             | HRV(mean NN)                           | 0.36 [0.13, 1.00]*                   |
| HRV(mean NN):PA(steps)                 | -0.13 [-0.39, 0.13]                            | HRV(SDNN)                              | 0.70 [0.31, 1.56]                    |
| HRV(SDNN)                              | -0.07 [-0.29, 0.15]                            | HRV(pNN20)                             | 1.04 [0.32, 3.36]                    |
| HRV(pNN20)                             | 0.38 [0.06, 0.71]*                             | HRV(SDNN):PA(steps)                    | 0.88 [0.33, 2.36]                    |
| HRV(pNN20):PA(steps)                   | 0.30 [-0.09, 0.70]                             | HRV(pNN20):PA(steps)                   | 1.43 [0.34, 5.99]                    |
| absolute jitter                        | 0.15 [-0.14, 0.45]                             | HRV(meanNN):PA(steps)                  | 0.97 [0.37, 2.53]                    |
| absolute jitter:PA(steps)              | -0.14 [-0.50, 0.21]                            | absolute jitter                        | 2.65 [0.75, 9.42]                    |
| absolute jitter:HRV(mean NN)           | 0.22 [-0.15, 0.58]                             | absolute jitter:HRV(pNN20)             | 0.86 [0.37, 1.98]                    |
| absolute jitter:HRV(mean NN):PA(steps) | 0.14 [-0.36, 0.63]                             | absolute jitter:HRV(mean NN):PA(steps) | 3.82 [1.10, 13.27]*                  |
| absolute jitter:HRV(SDNN)              | -0.06 [-0.39, 0.27]                            | apq3 shimmer                           | 1.39 [0.45, 4.27]                    |
| absolute jitter:HRV(SDNN):PA(steps)    | -0.07 [-0.48, 0.34]                            | apq3 shimmer:PA(steps)                 | 2.80 [0.51, 15.27]                   |
| absolute jitter:HRV(pNN20)             | -0.50 [-0.99, -0.01]*                          | apq3 shimmer:HRV(mean NN)              | 3.73 [0.99, 14.01]                   |
| absolute jitter:HRV(pNN20):PA(steps)   | -0.27 [-1.00, 0.46]                            | apq3 shimmer:HRV(SDNN)                 | 1.01 [0.25, 4.08]                    |
| apq3 shimmer                           | -0.22 [-0.53, 0.09]                            | apq3 shimmer:HRV(SDNN):PA(steps)       | 0.16 [0.03, 0.87]*                   |
| apq3 shimmer:PA(steps)                 | 0.20 [-0.21, 0.61]                             | apq11 shimmer                          | 0.76 [0.29, 1.98]                    |
| apq3 shimmer:HRV(mean NN)              | -0.25 [-0.63, 0.13]                            | apq11 shimmer:PA(steps)                | 2.74 [0.57, 13.06]                   |
| apq3 shimmer:HRV(mean NN):PA(steps)    | -0.21 [-0.77, 0.34]                            | apq11 shimmer:HRV(mean NN)             | 0.48 [0.18, 1.31]                    |
| apq3 shimmer:HRV(SDNN)                 | -0.06 [-0.45, 0.32]                            | apq11 shimmer:HRV(mean NN):PA(steps)   | 0.49 [0.11, 2.10]                    |
| apq3 shimmer:HRV(SDNN):PA(steps)       | -0.15 [-0.70, 0.40]                            | apq11 shimmer:HRV(SDNN)                | 1.36 [0.48, 3.84]                    |
| apq3 shimmer:HRV(pNN20)                | 0.53 [0.03, 1.02]*                             | apq11 shimmer:HRV(pNN20):PA(steps)     | 1.02 [0.31, 3.41]                    |
| apq3 shimmer:HRV(pNN20):PA(steps)      | 0.46 [-0.32, 1.24]                             | ddp jitter                             | 0.11 [0.02, 0.69]*                   |
| mean F2                                | -0.28 [-0.52, -0.04]*                          | ddp jitter:HRV(SDNN)                   | 0.53 [0.10, 2.78]                    |
| mean F2:PA(steps)                      | 0.12 [-0.24, 0.47]                             | ddp jitter:HRV(mean NN)                | 0.08 [0.02, 0.38]**                  |
| mean F2:HRV(mean NN)                   | -0.31 [-0.61, -0.01]*                          | ddp jitter:PA(steps)                   | 0.13 [0.01, 1.69]                    |
| mean F2:HRV(mean NN):PA(steps)         | -0.24 [-0.66, 0.19]                            | mean F0                                | 1.81 [0.56, 5.90]                    |
| mean F2:HRV(SDNN)                      | -0.06 [-0.33, 0.20]                            | mean F0:HRV(mean NN)                   | 0.23 [0.06, 0.93]*                   |
| mean F2:HRV(SDNN):PA(steps)            | -0.30 [-0.69, 0.08]                            | mean F0:HRV(SDNN)                      | 0.92 [0.21, 3.93]                    |
| mean F2:HRV(pNN20)                     | 0.03 [-0.45, 0.51]                             | mean F0:HRV(SDNN):PA(steps)            | 1.24 [0.26, 5.90]                    |
| mean F2:HRV(pNN20):PA(steps)           | 0.39 [-0.36, 1.13]                             | mean F0:HRV(pNN20)                     | 1.10 [0.22, 5.57]                    |
|  |  | mean F0:HRV(pNN20):PA(steps)           | 1.23 [0.23, 6.58]                    |
|  |  | mean F0:PA(steps)                      | 1.08 [0.18, 6.32]                    |
|  |  | std F0:HRV(mean NN)                    | 2.21 [0.54, 9.14]                    |
|  |  | std F0:HRV(mean NN):PA(steps)          | 1.49 [0.38, 5.91]                    |
|  |  | std F0:HRV(SDNN)                       | 0.47 [0.11, 2.02]                    |
|  |  | std F0:HRV(pNN20)                      | 1.40 [0.35, 5.61]                    |
|  |  | std F0:PA(steps)                       | 0.47 [0.05, 4.60]                    |

**Table 3.** The results of the multivariate linear mixed-effects models that were generated to estimate (a) daily symptom score and (b) likelihood of exacerbation. The models included significant speech features and their interaction effects with heart rate variability metrics and daily step count. Significant model results are indicated as follows: \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ . Refer to Table 6 for definitions of feature names and acronyms.

Table 2 shows the results of the linear mixed-effects models between speech features and either daily symptom score or exacerbations. There were no significant associations between syllabic nuclei features and either COPD outcome. Among phonation features, higher values of jitter and shimmer were associated with a decrease in daily symptom score (absolute jitter:  $\beta = -0.18$ ,  $p < .05$ , 95% CI [-0.34, -0.01]; apq3 shimmer:  $\beta = -0.21$ ,  $p < .05$ , 95% CI [-0.40, -0.01]) and lower odds of exacerbation (jitter: OR = 0.44,  $p < .01$ , 95% CI [0.25, 0.77]; shimmer: OR = 0.54,  $p < .01$ , 95% CI [0.37, 0.78]).

| HRV(pNN20) Value | < Mean - SD                                 | ∈ [Mean - SD, Mean] | ∈ [Mean, Mean + SD] | > Mean + SD            |
|------------------|---|---------------------|---------------------|------------------------|
|                  | Symptom Score $\beta$ -coefficient [95% CI] |                     |                     |                        |
| apq3 shimmer     | -   | 0.10 [-0.15, 0.34]  | -0.09 [-0.45, 0.27] | -0.61 [-1.07, -0.15]** |
| absolute jitter  | -   | 0.24 [0.01, 0.47]*  | -0.19 [-0.51, 0.14] | -0.54 [-0.92, -0.16]** |

**Table 4.** The results of the univariate linear mixed-effects models generated to estimate daily symptom score and stratified by different levels of HRV(pNN20) for speech features. Note that the column for HRV(pNN20) < Mean - SD is empty due to a lack of data in that quadrant. Significant model results are indicated as follows: \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ .

| HRV(mean NN) Value | < Mean - SD                      | ∈ [Mean - SD, Mean] | ∈ [Mean, Mean + SD] | > Mean + SD       |
|--------------------|----------------------------------|---------------------|---------------------|-------------------|
|                    | Exacerbation Odds Ratio [95% CI] |                     |                     |                   |
| ddp jitter         | 3.47 [3.12, 4.36]*               | 0.61 [0.29, 1.23]   | 0.07 [0.01, 0.37]** | 0.29 [0.08, 1.14] |
| mean F0            | 0.80 [0.19, 3.22]                | 2.53 [1.09, 5.85]*  | 2.32 [1.03, 5.20]*  | 1.05 [0.47, 2.34] |

**Table 5.** The results of the univariate linear mixed-effects models generated to estimate the likelihood of exacerbation and stratified by different levels of HRV(mean NN) for speech features. Significant model results are indicated as follows: \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ .

Among prosodic features, the mean value of the second formant showed a statistically significant negative association with daily symptom score (mean F2:  $\beta = -0.24$ ,  $p < .05$ , 95% CI [-0.48, -0.01]). Conversely, higher mean and standard deviation values of the fundamental frequency were associated with higher odds of exacerbation (mean F0: OR = 1.72,  $p < .01$ , 95% CI [1.12, 2.57]; std F0: OR = 1.62,  $p < .05$ , 95% CI [1.04, 2.53]).

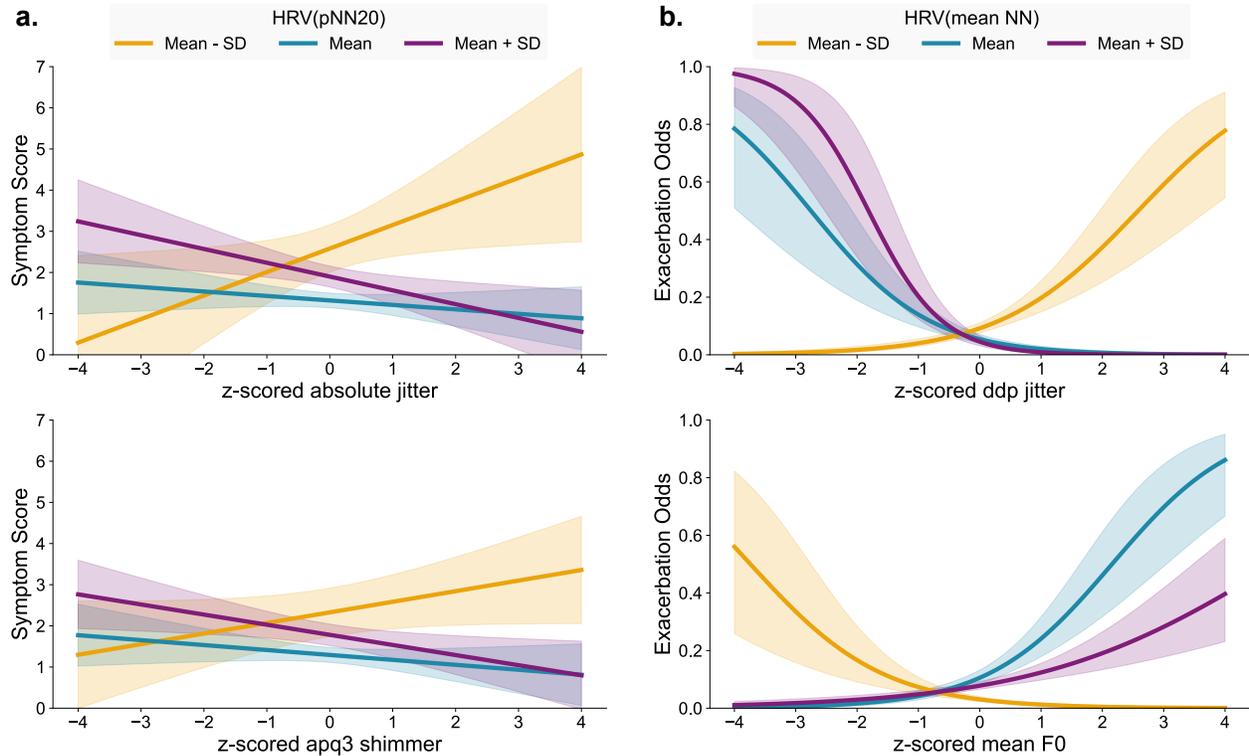
### 2.3 Multivariate associations between speech and COPD outcomes

Table 3a shows the results of the multivariate linear mixed-effects model between speech features and physiological covariates with respect to daily symptom score. This analysis only includes features that demonstrated significant univariate relationships without multicollinearity in relation to other features. The association with prosodic features was robust with respect to the physiological covariates. Similar to the corresponding univariate model, higher mean values of the second formant were associated with statistically significant decreases in daily symptom scores (mean F2:  $\beta = -0.24$ ,  $p < .05$ , 95% CI [-0.48, -0.01]). The same could not be said for the phonation features, jitter and shimmer, since their relationship with the symptom score lost its significance. The model also revealed significant interactions between heart rate variability (HRV) metrics and phonation features, including HRV(pNN20) with jitter, HRV(pNN20) with shimmer, and HRV(mean NN) with mean F2.

Since HRV(pNN20) emerged as a common moderator among the significant two-way interaction terms involving absolute jitter and apq3 shimmer in the symptom score model, we applied a moderation analysis to assess how varying levels of autonomic function according to HRV(pNN20) affected the direction and strength of association between speech features and COPD outcomes. The results of this analysis are shown in Table 4 and visualized in Figure 1(a). Both jitter and shimmer had a negative association with symptom score in the presence of high HRV(pNN20). However, an increase in jitter was associated with a statistically significant increase in symptom score when lower HRV(pNN20) values were measured.

Table 3b shows the results of the multivariate linear mixed-effects model between speech features and physiological covariates with respect to exacerbations. None of the speech features retained the statistical significance of their association. However, multiple two-way interactions across different categories of speech features were found to be significant: HRV(SDNN) with std F0, HRV(mean NN) with shimmer, HRV(pNN20) with shimmer, PA(steps) with jitter, and PA(steps) with mean F0. Table 5 illustrates the relationship between speech features and exacerbations when stratified by different levels of HRV(mean NN), as that was the most prominent HRV metric for this model. These significant interaction effects are visualized in Figure 1(b). This analysis allowed us to assess whether the strength and direction of these associations vary across levels of autonomic function, as measured by HRV(mean NN). The results show that lower values of jitter are associated with lower odds of exacerbation during periods with low HRV(mean NN). The opposite effect is observed in periods with higher HRV(mean NN), during which lower values of jitter are associated with an increased likelihood of exacerbation. Similarly, a positive association was observed between mean F0 and exacerbation during periods with moderate levels of HRV(mean NN).

Although no significant three-way interactions were detected for symptom scores, the fact that the same phonation features



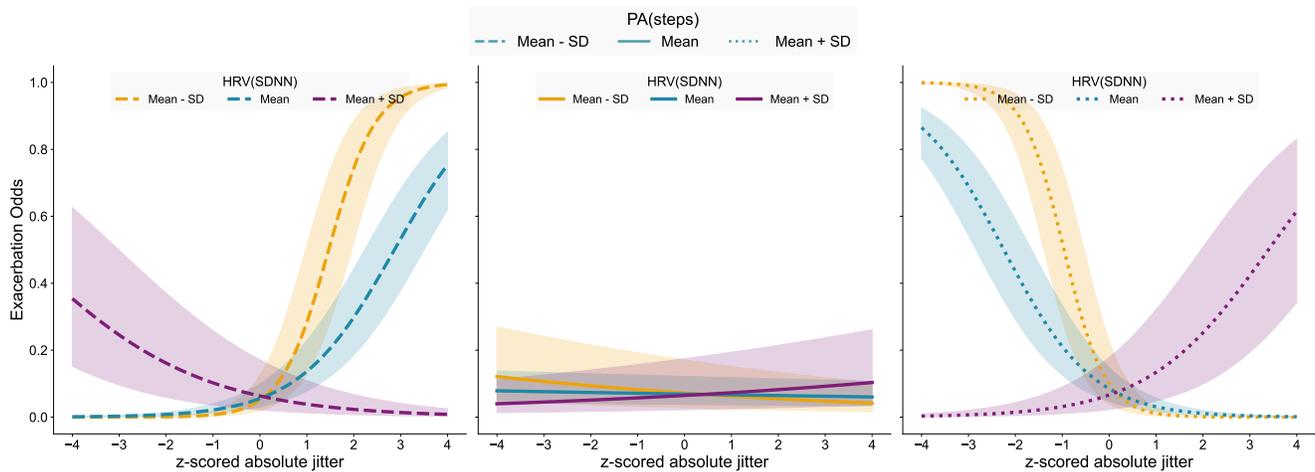
**Figure 1.** Interactions between HRV and z-score normalized speech features for (a) symptom score and (b) likelihood of exacerbation. The shaded regions surrounding each curve denote the 95% confidence intervals.

had significant two-way interactions with both HRV and physical activity indicates the importance of multimodal data. The interaction plots in Figure 2 depict the interplay between absolute jitter, HRV(SDNN), and PA(steps) with respect to the likelihood of exacerbations.

### 3 Discussion

When air is pushed through the vocal folds during speech production, it lends useful information about the functional limitations of people's respiratory airways. Although many researchers are turning towards deep learning to predict COPD outcomes from speech<sup>20,25,28</sup>, these techniques do not offer interpretable and explainable results that could inform future digital biomarkers. Others have examined manually extracted speech features as we do in our work<sup>19,24</sup>, often with a focus on overall predictive accuracy; however, there are a few that report on associations between speech features and outcomes. For example, Mohamed and el Maghraby<sup>17</sup> observed that jitter and shimmer were inversely correlated with spirometry metrics like FEV1 and FVC, while Farrús et al.<sup>18</sup> observed significantly slower speaking rates among COPD patients compared to healthy controls after physical activity.

Our study extends this literature by examining the associations between altered vocal characteristics and COPD outcomes, namely daily symptom score and exacerbations, during everyday living. We observed that categories of speech features reflecting different aspects of speech production exhibited unique relationships with COPD outcomes. Phonation features such as jitter, shimmer, and pitch and amplitude perturbations were found to have lower values when patients exhibited higher symptom scores or likelihood of an exacerbation. This trend may be attributed to increased vocal fold tension and rigidity due to heightened respiratory effort and stress<sup>29</sup>, resulting in a more stable but strained and higher-pitched voice. Our findings also reveal that prosodic features, specifically mean fundamental frequency (mean F0) and its standard deviation (std F0), were positively associated with exacerbations. On the other hand, formant frequencies like mean F2 showed a negative association with symptom scores. These observations diverge from previous research on acoustic measures for identifying pathological speech<sup>22,30</sup>. A key distinction in our study is the use of natural and unprompted speech, reflective of conditions in a remote patient monitoring setup. In free-living scenarios, patients may avoid speaking under extreme distress, so the observed vocal characteristics during this period likely stem from compensatory mechanisms where patients unconsciously adjust their vocal



**Figure 2.** The interaction between HRV(SDNN), PA(steps), and z-score normalized absolute jitter in speech with respect to exacerbations. The shaded regions surrounding each curve denote the 95% confidence intervals.

production to maintain intelligibility. Additionally, the negative association of mean F2 with symptom scores may result from reduced articulatory precision due to discomfort, leading to less distinct vowel articulation and lower F2 frequencies<sup>31</sup>.

We also extend the literature by showing that the relationships between speech features and COPD outcomes are influenced by physiological covariates inherent to everyday living, such as HRV and physical activity. Our interaction models revealed that during periods of higher HRV, there was a stronger association between lower jitter and worsened COPD outcomes; the same held true for lower shimmer. This suggests that patients with more adaptable cardiac autonomic function, as indicated by higher HRV, may experience a more stable vocal fold pattern, even under respiratory distress. However, when HRV is lower—a state linked to increased susceptibility to adverse COPD outcomes<sup>11</sup>—the relationship was reversed. In these instances, increased jitter correlated with higher symptom scores and a greater likelihood of exacerbations, consistent with findings from prior studies. This variation emphasizes the complex interplay between respiratory effort, autonomic regulation, and vocal production in COPD patients. The impact of physical activity on these interactions further highlights the vulnerability of patients with low activity levels. During periods of reduced activity, the association between both jitter and HRV with exacerbation risk became more pronounced, consistent with previous research identifying low physical activity as a marker of increased exacerbation risk<sup>32,33</sup>.

By analyzing the combined influence of speech characteristics, HRV, and physical activity on COPD outcomes, our study is the first to provide comprehensive insights into the physiological manifestation of COPD. Although the effect sizes of the interaction terms were typically modest, as expected in free-living conditions with real-world noise, the consistency and statistical significance of the associations highlight that physiological context matters when interpreting vocal markers. This multimodal approach not only provides a holistic view of COPD but also synthesizes and reconciles findings from previous studies that have individually linked various physiological parameters with COPD severity, thereby advancing the field with a unified perspective on disease pathophysiology. It paves the way for developing composite digital biomarkers that combine multiple data sources to predict adverse outcomes such as exacerbations. These digital biomarkers could provide real-time, actionable insights into a patient's health status, facilitating proactive management and personalized interventions. Additionally, our findings suggest that incorporating HRV metrics and step count into speech analysis models could potentially enhance the predictive power for symptom exacerbation and severity, enabling a more precise assessment of a patient's condition.

Our findings also underscore the potential of off-the-shelf wearables for continuous respiratory health monitoring. These devices could provide accessible, cost-effective alternatives to traditional lung function tests, which are often burdensome and limited in availability. The adoption of wearable technology in remote patient monitoring could enable patients to actively track their symptoms and severity while allowing clinicians to remotely observe and manage patient health. This continuous monitoring approach could enhance early detection of exacerbations, prompt timely medical interventions, and ultimately improve patient outcomes. Moreover, it has the potential to reduce healthcare costs by minimizing the need for frequent in-person visits and hospitalizations, making it a promising avenue for future research and development in COPD management.

We acknowledge several limitations in our study. Exacerbations are clinically defined according to self-reported symptoms, but the subjective nature of those self-reports may have implications on the robustness of our results. Although comparisons with objective pulmonary function tests (PFTs) would strengthen the validity of our results, prior literature has noted that lung function tests can be burdensome and inaccurate, particularly when done at home without supervision<sup>34-36</sup>. Using self-reported

symptoms yielded a more comprehensive and longitudinal assessment of the COPD condition, but future work could investigate accessible objective measures like at-home spirometry with automated coaching.

Another limitation of our study is the presence of noise in both speech and physiological data. We incorporated several measures to ensure quality in each stream of data. For speech, the SileroVAD voice activity detector<sup>37</sup> was used to exclude audio segments lacking discernible speech or containing excessive noise. For HRV, we relied on Samsung's proprietary heart rate tracking algorithm combined with filtering techniques to eliminate implausible inter-beat intervals and account for missing values. Similarly, Samsung's step-counting algorithm likely includes safeguards against misclassifying non-periodic movements as steps. Despite these measures, residual noise may have affected the accuracy of our measurements, so future studies could benefit from more advanced preprocessing techniques or alternative metrics that are robust to noise.

A further consideration is the potential influence of language on the observed relationships between speech and COPD outcomes. Although the acoustic features analyzed in this study primarily capture low-level aspects of speech production, cross-linguistic variation in phonemic structure, prosodic patterns, and speaking conventions may influence their distributions. Consequently, the strength or direction of the associations we uncovered may differ across languages. On a related note, the relatively small sample size for exacerbations may limit the generalizability of our findings. Our models accounted for demographic confounders to mitigate the influence of potential biases in the cohort, but we acknowledge that a larger and more diverse sample would improve the statistical significance and generalizability of our conclusions.

## 4 Methods

### 4.1 Subjects

A total of 35 patients (18 females, mean age:  $69.6 \pm 9.2$ ) diagnosed with COPD were recruited from three hospitals in Toronto, Canada to participate in our study. Participants were informed that they could request to withdraw from the study at any point in time. Those who required hospitalization or failed to adhere to the study protocol were de-enrolled. Study participants remained in the protocol for an average of  $198.9 \pm 122.6$  days. The analysis in this paper is restricted to those who were not de-enrolled and had more than 3 weeks of data, leaving us with 18 patients ranging in age from 55–93 years (6 females, mean age:  $67.3 \pm 8.5$  years).

### 4.2 Study design and data collection

The study protocol received approval from the Research Ethics Board at the University of Toronto under Protocol #41568. All participants provided informed consent, and the study was conducted in accordance with the Declaration of Helsinki. Participants collected data at home over the course of several months, with the total study period being approximately two years from the start date of the first participant to the end date of the last participant. Demographic data was collected at the start of the study, with participants reporting their sex, age, smoking history, and comorbidities in an electronic survey.

Participants were given a Samsung Galaxy Watch to collect data and a Samsung Galaxy Note smartphone with a pre-installed application to facilitate data transmission. Participants were asked to wear the watch during their daily activities and charge it at night, during which the smartphone application would identify and upload new data files. The smartwatch recorded audio data from the microphone at a sampling rate of 44.1 kHz, heart rate from the photoplethysmography (PPG) sensor at a sampling rate of 100 Hz, and daily step count from the motion data. To preserve the smartwatch's battery life, a 20% duty cycle was applied such that 2 minutes of data recording was followed by 8 minutes of idle time.

Participants also used the smartphone app to report the severity of their COPD symptoms every morning according to the London COPD Cohort Daily Symptom Questionnaire<sup>26,27</sup>, a clinically validated instrument that has been used in both clinical<sup>27,38</sup> and technical<sup>24,39</sup> studies to screen for COPD exacerbations. The questionnaire requires patients to reflect on the severity of the eight symptoms by selecting the ones that were "worse than usual on the previous day". The questionnaire considers three major symptoms—change in sputum color, increased breathlessness, and increased sputum amount—with a score of 5. The remaining symptoms—cold, fever, increased wheezing, sore throat, and worsening cough—are considered to be minor symptoms with a score of 1. An exacerbation is defined as two consecutive days with a symptom score of 6 or above, indicating two consecutive days with at least one major and one minor symptom. COPD symptom score and exacerbations constitute the two outcome variables investigated in this paper.

### 4.3 Data processing

As participants wore their smartwatches during their daily living, they may have produced non-speech sounds or encountered other ambient noises. To isolate useful speech segments from real-world audio data, we first removed intervals of silence by sliding a 50-ms window and discarding audio that fell below an overall intensity of -20 dB. The non-silent recordings were then split into 2-s windows with a 1-s step (50% overlap) and passed through a voice activity detector called SileroVAD<sup>37</sup>. SileroVAD is a neural network pre-trained on huge speech corpora including over 100 languages with various background noise and quality levels. The model takes a 20-ms audio frame to generate a speech probability score, which we then binarized using

| Data                                 | Category                               | Feature  | Description   |
|--------------------------------------|--|--|---|
| Speech                               | Phonation Features <sup>40</sup>       | HNR  | Harmonic sound to noise ratio   |
|                                      |  | jitter   | Frequency perturbation across vocal fold cycles   |
|                                      |  | absolute jitter  | Absolute frequency perturbation   |
|                                      |  | ddp jitter   | Average absolute difference between the frequency differences of adjacent pitch periods                     |
|                                      |  | ppq5 jitter  | Five-point perturbation quotient for vocal fold frequency   |
|                                      |  | rap jitter   | Relative average perturbation of vocal fold frequency   |
|                                      |  | shimmer  | Variability of the peak-to-peak amplitude   |
|                                      |  | apq3 shimmer   | Three-point amplitude perturbation quotient   |
|                                      |  | apq5 shimmer   | Five-point amplitude perturbation quotient  |
|                                      |  | apq11 shimmer  | Eleven-point amplitude perturbation quotient  |
|                                      | dda shimmer                            | Average absolute difference between the amplitudes of consecutive periods  |   |
|                                      | Prosodic Features <sup>41</sup>        | mean F0  | Mean fundamental frequency  |
|                                      |  | std F0   | Standard deviation of fundamental frequency   |
|                                      |  | mean F1  | Average value of the first formant frequency (F1)   |
|                                      |  | mean F2  | Average value of the second formant frequency (F2)  |
|                                      |  | mean F3  | Average value of the third formant frequency (F3)   |
|                                      |  | mean F4  | Average value of the fourth formant frequency (F4)  |
|                                      |  | mean formant   | Average value of all the four formants  |
|                                      |  | delta_f <sup>42</sup>  | Weighted average of the first four formant frequencies (F1, F2, F3, F4) based on their respective positions |
|                                      |  | fdisp <sup>43</sup>  | Formant dispersion, calculated as the average spacing between adjacent formant frequencies                  |
|                                      |  | fitch_vtl <sup>43</sup>  | Vocal tract length, calculated as the weighted average of the wavelengths of the first four formants        |
|                                      | mff <sup>44</sup>                      | Geometric mean of the first four formant frequencies   |   |
|                                      | pF <sup>45</sup>                       | Formant position, calculated as the average standardized deviation of the first four formant frequencies from their respective means |   |
|                                      | vtl_delta_f <sup>43</sup>              | Vocal tract length, calculated as the inverse of delta_f   |   |
|                                      | Syllabic Nuclei Features <sup>24</sup> | pause rate   | Total number of pauses divided by the duration of the segment   |
|                                      |  | phonation time   | Total duration of syllables divided by the voiced segment duration  |
|                                      |  | speech rate  | Total number of spoken syllables divided by the duration of the segment                                     |
| Heart Rate Variability <sup>46</sup> | HRV(mean NN)                           | Average of all the normal-to-normal (NN) intervals   |   |
|                                      | HRV(SDNN)                              | Standard deviation of all the NN intervals   |   |
|                                      | HRV(pNN20)                             | Proportion of successive NN intervals that differ by greater than 20 ms  |   |
| Physical Activity                    |  | PA(step count)   | Daily step count  |

**Table 6.** Overview of features used for analyzing sensor data.

an empirically determined threshold of 0.7 to separate speech and non-speech segments. The binary labels for each frame were aggregated back into their original 2-s windows to identify longer segments of contiguous speech. We discarded all utterances shorter than 250 ms based on the observation that most meaningful speech utterances are longer than that duration<sup>47</sup>. If the combined duration of all utterances in a window was longer than 1.5 s (75% of the window), we considered that window as speech and retained it for further analysis. Across participants, these processing steps yielded an average of  $122.6 \pm 97.4$  minutes of usable speech per day, from an initial average of  $161.3 \pm 123.2$  minutes per day of audio.

After identifying speech segments, we extracted three categories of speech features that are known to be correlated with health<sup>48,49</sup>: (1) phonation features, which capture the temporal and amplitude variations in the vibration of the vocal cords; (2) prosodic features, which measure the pitch and loudness of each frame; and (3) syllable nuclei features, which relate to the speaking rate according to syllables. Table 6 lists and defines all of the features we used to characterize the change in vocal characteristics of a patient. Because COPD symptom severity was self-reported daily in our dataset, we computed the daily average of each feature to generate a single feature vector summarizing a whole day of speech audio. This step also reduced short-term variability caused by transient factors such as background noise, yielding robust and stable features that better reflect the gradual changes patients typically exhibit over daily timescales<sup>26</sup>.

We also examined the interplay between physical activity and heart rate variability (HRV) as physiological covariates with speech to characterize vocal changes and COPD severity more comprehensively. For physical activity, we used the aggregate daily step count calculated by Samsung's proprietary algorithm; we call this covariate PA(steps). Daily step count data were available for an average of  $136.7 \pm 86.4$  days per participant. For HRV, we first processed the 1 Hz heart rate data reported by Samsung's proprietary algorithm. PPG data collected in the wild contains noise from several sources, including motion artifacts, wristband looseness, and environmental light<sup>50</sup>. As a result, the heart rate information provided by the smartwatch may have missing values that can impact derived features. We only considered heart rate data segments with a minimum length of 45 s in order to preserve the low-frequency component of the HRV signal (0.04–0.15 Hz)<sup>51</sup>. To calculate HRV, the heart rate data was converted to RR intervals according to the hyperbolic relationship between the two ( $HR \times RR \text{ interval} = 60,000$ )<sup>52</sup>. We then used a range-based filter to remove physiologically impossible values below 350 ms and above 1,200 ms. Finally, we calculated three HRV metrics informed by prior literature<sup>53,54</sup> in each segment: HRV(mean NN), the mean of normal RR intervals; HRV(SDNN), the standard deviation of normal RR intervals; and HRV(pNN20), the percentage of consecutive normal RR intervals that differed by more than 20 ms. Similar to speech and activity data, these metrics were averaged across a day to represent daily HRV. Participants had an average of  $85.5 \pm 67.0$  days of usable HRV data, compared to  $138.1 \pm 87.7$  days prior to data processing.

#### 4.4 Statistics

Given the large number of candidate speech features and interaction terms, we adopted a two-step procedure to identify a subset appropriate for inclusion in the multivariate models alongside HRV and physical activity data. First, we fit univariate linear mixed-effects models for all speech features, with either COPD symptom score or exacerbation as the outcome variable. Next, we calculated variance inflation factor (VIF) values to determine whether the inclusion of the significant features and their corresponding interaction terms would create multicollinearity in the ultimate multivariate model. Starting with the full model including all of the terms, we iteratively removed those with the highest VIF values and re-estimated the model until all retained terms had VIFs were below 20.

Once we had our final set of statistically significant and uncorrelated speech features, we generated a multivariate model that included those features along with our four physiological covariates—three HRV metrics and daily step count—as modifiers of vocal changes associated with COPD outcomes. The multivariate model also included two-way interactions between the significant speech features and each of the physiological covariates, along with three-way interactions involving all three data types. For physiological covariates that appeared in multiple significant interaction terms, we conducted a moderation analysis inspired by Bauer and Curran<sup>55</sup> to examine whether the direction or strength of association between speech features and outcomes was affected by different covariate levels. We first stratified the data according to whether the notable physiological covariate fell within one of four intervals:  $< \text{Mean} - \text{SD}$ ,  $[\text{Mean} - \text{SD}, \text{Mean}]$ ,  $[\text{Mean}, \text{Mean} + \text{SD}]$ , and  $> \text{Mean} + \text{SD}$ . We then fitted separate univariate models for the corresponding speech features in the significant interaction terms.

All analyses were adjusted for three demographic confounders selected from prior literature<sup>56</sup>: age, sex, and smoking history. We report the  $\beta$ -coefficients for symptom score relationships and odds ratios for exacerbation relationships. We also include the 95% confidence intervals (CI) and p-values for these relationships with statistical significance being set at  $p < .05$ . All analyses were conducted in R-3.5.1 using the `lmer`<sup>57</sup>, `tidyverse`<sup>58</sup>, and `interactions`<sup>59</sup> packages.

#### Data availability

On reasonable request, the dataset generated during and/or analyzed in this paper is available from the corresponding author.

## Code availability

The code used for the analysis of the dataset is available upon reasonable request from the corresponding author.

## Acknowledgements

The Authors would like to acknowledge Teresa To for her feedback on the statistical analysis presented in this work. This study was funded in part by Canadian Institutes of Health Research (CIHR) Grant #155458, Natural Sciences and Engineering Research Council of Canada (NSERC) Grants #RGPIN-2021-03457 and #RGPIN-2017-06618, and Samsung Research America (SRA). This was an investigator-initiated study with input from SRA. The funders played no role in the analysis and interpretation of data, or the writing of this manuscript.

## Author contributions

S.L., D.L., R.W., A.G., and E.L. conceived and executed the initial study that produced the dataset. S.B., R.W., A.G., E.L., and A.M. conceived the experiments for this paper. S.B. and D.K. analysed the data. S.B. wrote the main manuscript text. All authors reviewed the manuscript.

## Competing Interests

The Authors declare no competing financial or non-financial interests.

## References

1. Momtazmanesh, S. *et al.* Global burden of chronic respiratory diseases and risk factors, 1990–2019: an update from the Global Burden of Disease Study 2019. *eClinicalMedicine* **59**, 101936, DOI: [10.1016/j.eclinm.2023.101936](https://doi.org/10.1016/j.eclinm.2023.101936) (2023).
2. Roche, N. *et al.* COPD exacerbations significantly impact quality of life as measured by SGRQ-C total score: results from the FLAME study. *Eur. Respir. J.* **50**, DOI: [10.1183/1393003.congress-2017.OA1487](https://doi.org/10.1183/1393003.congress-2017.OA1487) (2017).
3. Troosters, T. *et al.* Physical inactivity in patients with COPD, a controlled multi-center pilot-study. *Respir. Medicine* **104**, 1005–1011, DOI: <https://doi.org/10.1016/j.rmed.2010.01.012> (2010).
4. Watz, H., Waschki, B., Meyer, T. & Magnussen, H. Physical activity in patients with COPD. *Eur. Respir. J.* **33**, 262–272, DOI: [10.1183/09031936.00024608](https://doi.org/10.1183/09031936.00024608) (2009).
5. Suissa, S., Dell’Aniello, S. & Ernst, P. Long-term natural history of chronic obstructive pulmonary disease: severe exacerbations and mortality. *Thorax* **67**, 957–963, DOI: [10.1136/thoraxjnl-2011-201518](https://doi.org/10.1136/thoraxjnl-2011-201518) (2012).
6. Jeetvan G Patel, O. E. L., Anna D Coutinho & Dalal, A. A. COPD affects worker productivity and health care costs. *Int. J. Chronic Obstr. Pulm. Dis.* **13**, 2301–2311, DOI: [10.2147/COPD.S163795](https://doi.org/10.2147/COPD.S163795) (2018). PMID: 30104870.
7. Banerji, D., Kulich, K., Keininger, D. L. & Tiplady, B. Symptoms and impact of COPD assessed by an electronic diary in patients with moderate-to-severe COPD: psychometric results from the SHINE study. *Int. J. Chronic Obstr. Pulm. Dis.* **79**, DOI: [10.2147/COPD.S73092](https://doi.org/10.2147/COPD.S73092) (2015).
8. Johns, D. P., Walters, J. A. E. & Walters, E. H. Diagnosis and early detection of COPD using spirometry. *J. Thorac. Dis.* **6**, 1557–1569, DOI: [10.3978/j.issn.2072-1439.2014.08.18](https://doi.org/10.3978/j.issn.2072-1439.2014.08.18) (2014).
9. Johnston, N. W. *et al.* Detection of COPD Exacerbations and Compliance With Patient-Reported Daily Symptom Diaries Using a Smartphone-Based Information System. *Chest* **144**, 507–514, DOI: [10.1378/chest.12-2308](https://doi.org/10.1378/chest.12-2308) (2013).
10. Wood-Baker, R., Walters, Robinson, A., Walters & Wills. Clinical diaries in COPD: compliance and utility in predicting acute exacerbations. *Int. J. Chronic Obstr. Pulm. Dis.* **427**, DOI: [10.2147/COPD.S32222](https://doi.org/10.2147/COPD.S32222) (2012).
11. Alqahtani, J. S. *et al.* A systematic review and meta-analysis of heart rate variability in COPD. *Front. Cardiovasc. Medicine* **10**, 1070327, DOI: [10.3389/fcvm.2023.1070327](https://doi.org/10.3389/fcvm.2023.1070327) (2023).
12. Bernardo, A. F. B., Vanderlei, L. C. M. & Garner, D. M. HRV Analysis: A Clinical and Diagnostic Tool in Chronic Obstructive Pulmonary Disease. *Int. Sch. Res. Notices* **2014**, 1–6, DOI: [10.1155/2014/673232](https://doi.org/10.1155/2014/673232) (2014).
13. Liao, S.-Y., Benzo, R., Ries, A. L. & Soler, X. Physical Activity Monitoring in Patients with Chronic Obstructive Pulmonary Disease. *Chronic Obstr. Pulm. Dis. (Miami, Fla.)* **1**, 155–165, DOI: [10.15326/jcopdf.1.2.2014.0131](https://doi.org/10.15326/jcopdf.1.2.2014.0131) (2014).
14. Pericleous, P. & van Staa, T. P. The use of wearable technology to monitor physical activity in patients with COPD: a literature review. *Int. J. Chronic Obstr. Pulm. Dis.* **14**, 1317–1322, DOI: [10.2147/COPD.S193037](https://doi.org/10.2147/COPD.S193037) (2019).

15. Vyshedskiy, A. & Murphy, R. Acoustic biomarkers of Chronic Obstructive Lung Disease. *Res. Ideas Outcomes* **2**, e9173, DOI: [10.3897/rio.2.e9173](https://doi.org/10.3897/rio.2.e9173) (2016).
16. Mohammed Daanish Ali Khan, M., Pradeep Naval, P., Kulshreshtha, R., Venneti, S. & Singh, A. VOICE-BASED MONITORING OF COPD. *Chest* **160**, A2173–A2174, DOI: [10.1016/j.chest.2021.07.1920](https://doi.org/10.1016/j.chest.2021.07.1920) (2021).
17. Mohamed, E. E. & El Maghraby, R. A. Voice changes in patients with chronic obstructive pulmonary disease. *Egypt. J. Chest Dis. Tuberc.* **63**, 561–567, DOI: <https://doi.org/10.1016/j.ejcdt.2014.03.006> (2014).
18. Farrús, M. *et al.* Speech-Based Support System to Supervise Chronic Obstructive Pulmonary Disease Patient Status. *Appl. Sci.* **11**, 7999, DOI: [10.3390/app11177999](https://doi.org/10.3390/app11177999) (2021).
19. Chun, K. S. *et al.* Towards passive assessment of pulmonary function from natural speech recorded using a mobile phone. In *2020 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 1–10, DOI: [10.1109/PerCom45495.2020.9127380](https://doi.org/10.1109/PerCom45495.2020.9127380) (2020).
20. Nallanthighal, V. S., Härmä, A. & Strik, H. Detection of copd exacerbation from speech: comparison of acoustic features and deep learning based speech breathing models. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 9097–9101 (IEEE, 2022).
21. Saleheen, N. *et al.* Lung Function Estimation from a Monosyllabic Voice Segment Captured Using Smartphones. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*, 1–11, DOI: [10.1145/3379503.3403543](https://doi.org/10.1145/3379503.3403543) (ACM, Oldenburg Germany, 2020).
22. Merkus, J., Hubers, F., Cucchiarini, C. & Strik, H. Digital Eavesdropper-acoustic speech characteristics as markers of exacerbations in COPD patients. In *Proceedings of RaPID Workshop of the 12th International Conference on Language Resources and Evaluation (LREC2020)* (2020).
23. Nathan, V., Vatanparvar, K., Rahman, M. M., Nemati, E. & Kuang, J. Assessment of Chronic Pulmonary Disease Patients Using Biomarkers from Natural Speech Recorded by Mobile Devices. In *2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 1–4, DOI: [10.1109/BSN.2019.8771043](https://doi.org/10.1109/BSN.2019.8771043) (IEEE, Chicago, IL, USA, 2019).
24. Sedaghat, T. *et al.* Unobtrusive Monitoring of COPD Patients using Speech Collected from Smartwatches in the Wild. In *2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*, 818–823, DOI: [10.1109/PerComWorkshops53856.2022.9767283](https://doi.org/10.1109/PerComWorkshops53856.2022.9767283) (IEEE, Pisa, Italy, 2022).
25. Bhalla, S. *et al.* Pulmolistener: Continuous acoustic monitoring of chronic obstructive pulmonary disease in the wild. *Proc. ACM on Interactive, Mobile, Wearable Ubiquitous Technol.* **7**, 1–24 (2023).
26. Aaron, S. D. *et al.* Time course and pattern of COPD exacerbation onset. *Thorax* **67**, 238–243, DOI: [10.1136/thoraxjnl-2011-200768](https://doi.org/10.1136/thoraxjnl-2011-200768) (2012).
27. Seemungal, T. A. R., Donaldson, G. C., Bhowmik, A., Jeffries, D. J. & Wedzicha, J. A. Time Course and Recovery of Exacerbations in Patients with Chronic Obstructive Pulmonary Disease. *Am. J. Respir. Critical Care Medicine* **161**, 1608–1613, DOI: [10.1164/ajrccm.161.5.9908022](https://doi.org/10.1164/ajrccm.161.5.9908022) (2000).
28. Srivastava, A. *et al.* Deep learning based respiratory sound analysis for detection of chronic obstructive pulmonary disease. *PeerJ Comput. Sci.* **7**, e369 (2021).
29. Yao, X., Jitsuhiro, T., Miyajima, C., Kitaoka, N. & Takeda, K. Physical characteristics of vocal folds during speech under stress. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4609–4612 (IEEE, 2012).
30. Shastry, A., Balasubramaniam, R. K. & Acharya, P. R. Voice Analysis in Individuals with Chronic Obstructive Pulmonary Disease. *Int. J. Phonosurgery & Laryngol.* **4**, 45–49, DOI: [10.5005/jp-journals-10023-1081](https://doi.org/10.5005/jp-journals-10023-1081) (2014).
31. Dejonckere, P. H. Assessment of voice and respiratory function. In *Surgery of larynx and trachea*, 11–26 (Springer, 2009).
32. Garcia-Rio, F. *et al.* Prognostic value of the objective measurement of daily physical activity in patients with COPD. *Chest* **142**, 338–346, DOI: <https://doi.org/10.1378/chest.11-2014> (2012).
33. Pitta, F. *et al.* Physical activity and hospitalization for exacerbation of copd. *Chest* **129**, 536–544, DOI: <https://doi.org/10.1378/chest.129.3.536> (2006).
34. Anand, R. *et al.* Unsupervised home spirometry versus supervised clinic spirometry for respiratory disease: a systematic methodology review and meta-analysis. *Eur. Respir. Rev.* **32**, 220248, DOI: [10.1183/16000617.0248-2022](https://doi.org/10.1183/16000617.0248-2022) (2023).

35. Moor, C. C. *et al.* A home monitoring program including real-time wireless home spirometry in idiopathic pulmonary fibrosis: a pilot study on experiences and barriers. *Respir. Res.* **19**, 105, DOI: [10.1186/s12931-018-0810-3](https://doi.org/10.1186/s12931-018-0810-3) (2018).
36. Talboom-Kamp, E. P., Holstege, M. S., Chavannes, N. H. & Kasteleyn, M. J. Effects of use of an eHealth platform e-Vita for COPD patients on disease specific quality of life domains. *Respir. Res.* **20**, 146, DOI: [10.1186/s12931-019-1110-2](https://doi.org/10.1186/s12931-019-1110-2) (2019).
37. SileroVAD: pre-trained enterprise-grade voice activity detector (VAD), number detector and language classifier. <https://github.com/snakers4/silero-vad> (2021).
38. Seemungal, T. A. *et al.* Effect of exacerbation on quality of life in patients with chronic obstructive pulmonary disease. *Am. journal respiratory critical care medicine* **157**, 1418–1422 (1998).
39. Liaqat, D. *et al.* Coughwatch: Real-World Cough Detection using Smartwatches. In *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 8333–8337, DOI: [10.1109/ICASSP39728.2021.9414881](https://doi.org/10.1109/ICASSP39728.2021.9414881) (IEEE, Toronto, ON, Canada, 2021).
40. Teixeira, J. P., Oliveira, C. & Lopes, C. Vocal Acoustic Analysis – Jitter, Shimmer and HNR Parameters. *Procedia Technol.* **9**, 1112–1122, DOI: [10.1016/j.protcy.2013.12.124](https://doi.org/10.1016/j.protcy.2013.12.124) (2013).
41. Kent, R. D. & Vorperian, H. K. Static measurements of vowel formant frequencies and bandwidths: A review. *J. Commun. Disord.* **74**, 74–97, DOI: [10.1016/j.jcomdis.2018.05.004](https://doi.org/10.1016/j.jcomdis.2018.05.004) (2018).
42. Reby, D. & McComb, K. Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. *Animal Behav.* **65**, 519–530, DOI: [10.1006/anbe.2003.2078](https://doi.org/10.1006/anbe.2003.2078) (2003).
43. Fitch, W. T. Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The J. Acoust. Soc. Am.* **102**, 1213–1222, DOI: [10.1121/1.421048](https://doi.org/10.1121/1.421048) (1997).
44. Smith, D. R. R. & Patterson, R. D. The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *The J. Acoust. Soc. Am.* **118**, 3177–3186, DOI: [10.1121/1.2047107](https://doi.org/10.1121/1.2047107) (2005).
45. Puts, D. A., Apicella, C. L. & Cárdenas, R. A. Masculine voices signal men’s threat potential in forager and industrial societies. *Proc. Royal Soc. B: Biol. Sci.* **279**, 601–609, DOI: [10.1098/rspb.2011.0829](https://doi.org/10.1098/rspb.2011.0829) (2012).
46. Malik, M. *et al.* Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *Eur. Hear. J.* **17**, 354–381, DOI: [10.1093/oxfordjournals.eurheartj.a014868](https://doi.org/10.1093/oxfordjournals.eurheartj.a014868) (1996).
47. Corps, R. E., Knudsen, B. & Meyer, A. S. Overrated gaps: Inter-speaker gaps provide limited information about the timing of turns in conversation. *Cognition* **223**, 105037, DOI: <https://doi.org/10.1016/j.cognition.2022.105037> (2022).
48. Bayestehtashk, A., Asgari, M., Shafran, I. & McNamers, J. Fully automated assessment of the severity of Parkinson’s disease from speech. *Comput. Speech & Lang.* **29**, 172–185, DOI: <https://doi.org/10.1016/j.csl.2013.12.001> (2015).
49. Lee, K., Moon, C. & Nam, Y. Diagnosing vocal disorders using cobweb clustering of the jitter, shimmer, and harmonics-to-noise ratio. *KSII Transactions on Internet Inf. Syst.* **12**, 5541–5554, DOI: [10.3837/tiis.2018.11.020](https://doi.org/10.3837/tiis.2018.11.020) (2018).
50. Ghamari, M., Nazeran, H. & Soltanpur, C. A review on wearable photoplethysmography sensors and their potential future applications in health care. *Int. J. Biosens. & Bioelectron.* **4**, 195–202, DOI: [10.15406/ijbsbe.2018.04.00125](https://doi.org/10.15406/ijbsbe.2018.04.00125) (2018).
51. Tiwari, A. *et al.* Prediction of stress and mental workload using ultra-short-term heart rate variability and breathing analysis. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 4530–4533, DOI: [10.1109/EMBC44109.2020.9175414](https://doi.org/10.1109/EMBC44109.2020.9175414) (2020).
52. Young, D. W. Self-measure of heart rate variability (HRV) and arrhythmia to monitor and to manage atrial arrhythmias: personal experience with high intensity interval exercise (HIIE) for the conversion to sinus rhythm. *Front. Physiol.* **5**, DOI: [10.3389/fphys.2014.00251](https://doi.org/10.3389/fphys.2014.00251) (2014).
53. Ma, Y., Chang, M.-C., Litrownik, D., Wayne, P. M. & Yeh, G. Y. Day–night patterns in heart rate variability and complexity: differences with age and cardiopulmonary disease. *J. Clin. Sleep Medicine* **19**, 873–882 (2023).
54. Tiwari, A. *et al.* Remote copd severity and exacerbation detection using heart rate and activity data measured from a wearable device. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 7450–7454 (IEEE, 2021).
55. Bauer, D. J. & Curran, P. J. Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivar. Behav. Res.* **40**, 373–400, DOI: [10.1207/s15327906mbr4003\\_5](https://doi.org/10.1207/s15327906mbr4003_5) (2005). PMID: 26794689, [https://doi.org/10.1207/s15327906mbr4003\\_5](https://doi.org/10.1207/s15327906mbr4003_5).

56. Terzikhan, N. *et al.* Prevalence and incidence of COPD in smokers and non-smokers: the Rotterdam Study. *Eur. J. Epidemiol.* **31**, 785–792, DOI: [10.1007/s10654-016-0132-z](https://doi.org/10.1007/s10654-016-0132-z) (2016).
57. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4 (2014).
58. Wickham, H. *et al.* Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
59. Long, J. A. *Interactions: Comprehensive, User-Friendly Toolkit for Probing Interactions*, DOI: [10.32614/CRAN.package.interactions](https://doi.org/10.32614/CRAN.package.interactions) (2024). R package version 1.2.0.

## Figure Legend

**Figure 1.** Interaction effects between HRV and standardized speech features. (a) Interaction effects between HRV(pNN20) and two phonation features, absolute jitter and APQ3 shimmer, on daily symptom score. (b) Interaction effects between HRV(mean NN) and two speech features, ddp jitter and mean F0, on the likelihood of an exacerbation. Speech feature values on the x-axis are z-score normalized (mean = 0, SD = 1), with the plotted range spanning  $\pm 4$  standard deviations to capture the distribution of standardized feature values. Shaded regions indicate 95% confidence intervals around model estimates. Each curve shows the predicted association between the speech feature and outcome at different levels of HRV, highlighting how autonomic regulation moderates the relationship between speech and COPD status.

**Figure 2.** Three-way interaction between HRV, physical activity, and speech on exacerbation odds. The plot shows predicted exacerbation risk across levels of absolute jitter, stratified by HRV(SDNN) and PA(steps). Speech feature values on the x-axis are standardized (mean = 0, SD = 1), and the shaded regions represent 95% confidence intervals around model predictions. The visualization illustrates how the association between vocal stability and exacerbation risk varies with autonomic function and physical activity levels.

## Table Legend

**Table 1.** Participant characteristics and vital signs of COPD patients. Values are reported as median (interquartile range, IQR) or mean (standard deviation, SD), as appropriate.

**Table 2.** Univariate associations between speech features and COPD outcomes. Results are from linear mixed-effects models evaluating the relationship between free-living speech characteristics and COPD outcomes. Significant model results are indicated as follows: \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ . Refer to Table 6 for definitions of feature names and acronyms.

**Table 3.** Multivariate associations between speech features, physiological covariates, and COPD outcomes. Results are from linear mixed-effects models estimating (a) daily symptom score and (b) likelihood of exacerbation. Models include significant speech features from univariate analysis, along with their interaction terms with heart rate variability metrics and daily step count. Significant model results are indicated as follows: \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ . Refer to Table 6 for definitions of feature names and acronyms.

**Table 4.** Stratified associations between speech features and daily symptom score across different levels of HRV(pNN20). Results are from univariate linear mixed-effects models stratified by HRV(pNN20) to examine how associations between speech features and symptom score vary across different levels of heart rate variability. The column corresponding to HRV(pNN20) < Mean - SD is empty due to insufficient data in that range. Significant model results are indicated as follows: \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ .

**Table 5.** Stratified associations between speech features and exacerbation likelihood across different levels of HRV(mean NN). Results are from univariate linear mixed-effects models stratified by HRV(mean NN) to assess whether associations between speech features and exacerbation risk vary across different levels of heart rate variability. Significant model results are indicated as follows: \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ .

**Table 6.** Overview of features used in the analysis of speech and physiological data. This table defines the acoustic and physiological features included in the univariate and multivariate models.