Supplemental Material: The Impact of Symptom and Activity Trade-offs on Transmission Potential of Patients Infected with Influenza

2020-04-13

Table of Contents

# Authors

* Brian McKay1, Department of Epidemiology and Biostatistics, The University of Georgia, Athens, GA, USA
* Mark Ebell, Department of Epidemiology and Biostatistics, The University of Georgia, Athens, GA, USA
* Ariella Perry, Colorado Department of Public Health and Environment, Denver, CO, USA
* Ye Shen, Department of Epidemiology and Biostatistics, The University of Georgia, Athens, GA, USA
* Andreas Handel1, Department of Epidemiology and Biostatistics, The University of Georgia, Athens, GA, USA

1 Corresponding Authors: Brian McKay and Andreas Handel

Address: 101 Buck Rd, Miller Hall, Athens, Georgia 30606

Email: [bmckay52@uga.edu](mailto:bmckay52@uga.edu) or [ahandel@uga.edu](mailto:ahandel@uga.edu)

PROCEEDINGS OF THE ROYAL SOCIETY B  
DOI: 10.1098/rspb.2020.0496

# Documentation for additional SM files

## Getting the files

The files described below are available from Dryad <https://doi.org/10.5061/dryad.51c59zw4v>.

* “**Virulence\_Trade-off.Rproj**” This file lets R know the relative file paths for loading and saving files.
* “**SymptomActivity.bib**” This file has all of the citation saved as a bibTex.
* “**Symptom Questionnaire\_Redacted.pdf**”: This is a copy of the electronic questionnaire patients with an upper respiratory symptoms were required to fill out. All identifying information has been redacted.
* “**DataDictionary\_VirulenceTradeOff.xlsx**”: This document provides a description of all the variables included in the analysis.
* “**1 Anonymized Data**” folder contains the de-identified data for the analyses
  + An R script that merges all of the individual data sets and creates and saves “Data.Rda” and “Data.csv” in the “1 Anonymized Data” Folder. This script is not included since the raw data sets are not included to protect patient privacy.
* “**2 Data Cleaning Script**” folder contains one R script that cleans the data for the analyses
  + “**Data Merging Script.R**”: R script that merges all of the individual data sets and creates and saves “Data.Rda” in the “1 Anonymized Data” Folder. This script is not included since the raw data sets are not included to protect patient privacy.
  + “**Data Cleaning.R**”: This R script does all of the data preparation, creating all the required variables for the analysis. This script also produces the data sets used for the analyses and saves them in the “3 Clean Data” folder.
* “**4 Analysis Scripts**” folder has 4 R scripts that analyze the clean data and produce the results presented in the main text and supplement.
  + “**Flu Symptoms Activity Models.R**” This script creates the univariate/multivariate linear regression table, the Spearman rank correlation, and the CMH trend tests. Results are saved in “5 Results”
  + “**Flu Symptoms Activity Plots.R**” This script creates all of the plots. Results are saved in “5 Results”
  + “**Flu Symptoms Activity Tables.R**” This script creates all of the tables. Results are saved in “5 Results”
  + “**Multivariate Subset Selection.R**” This script does the variable selection for the multivariate model. Results are saved in “5 Results”
* “**6 Manuscript**” folder has 2 files in it used to create the manuscript.
  + “**Manuscript.Rmd**” This Rmd file creates the basic manuscript word document (formatting will not be identical)
  + “**proceedings-of-the-royal-society-b.csl**” is a style file to format the citations in the manuscript
* “**7 Supplemental Material**” folder has 2 files used to create this document
  + “**Supplemental Material.Rmd**” This Rmd file creates the basic supplemental material word document (formatting will not be identical)
  + “**proceedings-of-the-royal-society-b.csl**” is a style file to format the citations in the supplement

## Reproducing results

The files required to reproduce the results are 2 R Markdown files, 4 R script, and one anonymized data file. These files allow the reproduction of all results shown in the main text and SM. To reproduce the results follow these steps.

First, install R, Rstudio, and Pandoc (when you install Rstudio Pandoc should automatically install). Microsoft Word or Open Office Word is also required.

Second, save the zip file from Dryad on your local computer. Open the folder and double click “Virulence\_Trade-off”. This should open Rstudio (if prompted, select Rstudio as the app to open this file type). Then open and run the files below in the specified order.

1. R script “Data Cleaning.R” in the “2 Data Cleaning Script” folder uses “Data.Rda” and produces two clean data sets used for all further analyses. The data sets are all saved in the “3 Clean Data” folder and include:
   1. “SympAct\_Any\_Pos.Rda” Contains data for all influenza patients regardless of diagnosis method.
   2. “SympAct\_Lab\_Pos.Rda” Contains data for influenza patients diagnosed based on a PCR or rapid antigen test.

**It is important to note that “SympAct\_Lab\_Pos.Rda” is a subset of “SympAct\_Any\_Pos.Rda” based on the method of diagnosis.**

1. Four R scripts in the “4 Analysis Scripts” folder (“Flu Symptoms Activity Univariate Model.R”, “Flu Symptoms Activity Univariate Plots.R”, and “Flu Symptoms Activity Univariate Tables.R”, “Multivariate Subset Selection.R”). The order you run these scripts does not matter. Results of each script are automatically saved in the “5 Results” folder.
2. R Markdown “Manuscript.Rmd” is in the “6 Manuscript” folder. This combines all the relevant results and creates the main text as a Word document (some reformatting is required).
3. R Markdown “Supplemental Material.Rmd” in the “7 Supplemental Material” folder generates the supplementary material as Word document.

# Supplementary Results

## Histogram of reported activity levels

Reported activity levels ranging from 0 to 10 with a median of 4 for those patients with a lab diagnosis of influenza (SM Figure 1)

Figure 1: Histogram of reported activity levels for patients with a lab diagnosis of influenza.

Figure 1: Histogram of reported activity levels for patients with a lab diagnosis of influenza.

## Table of symptoms

SM Table 1 shows the symptoms among patients with a lab-based diagnosis.

Table 1: **Symptoms of the 324 patients with laboratory based flu diagnosis. The table shows the number of patients who reported having the following symptoms and the corresponding percentage.**

|  |  |
| --- | --- |
|  | Overall |
| n | 324 |
| Abdominal Pain = Yes (%) | 38 (11.7) |
| Breathlessness = Yes (%) | 128 (39.5) |
| Chest Congestion = Yes (%) | 194 (59.9) |
| Chest Pain = Yes (%) | 110 (34.0) |
| Chills/Sweats = Yes (%) | 287 (88.6) |
| Cough = Yes (%) | 306 (94.4) |
| Diarrhea = Yes (%) | 40 (12.3) |
| Ear Pain = Yes (%) | 59 (18.2) |
| Eye Pain = Yes (%) | 48 (14.8) |
| Fatigue = Yes (%) | 304 (93.8) |
| Headache = Yes (%) | 273 (84.3) |
| Itchy Eyes = Yes (%) | 75 (23.1) |
| Myalgia = Yes (%) | 290 (89.5) |
| Nasal Congestion = Yes (%) | 255 (78.7) |
| Nausea = Yes (%) | 119 (36.7) |
| Runny Nose = Yes (%) | 234 (72.2) |
| Sleeplessness = Yes (%) | 183 (56.5) |
| Sneeze = Yes (%) | 177 (54.6) |
| Sore Throat = Yes (%) | 265 (81.8) |
| Subjective Fever = Yes (%) | 242 (74.7) |
| Swollen Lymph Nodes = Yes (%) | 127 (39.2) |
| Tooth Pain = Yes (%) | 60 (18.5) |
| Vomiting = Yes (%) | 43 (13.3) |
| Weakness = Yes (%) | 306 (94.4) |
| Wheezing = Yes (%) | 105 (32.4) |

## Univariate and subset selection

Correlations between activity level and each symptom from the univariate linear analysis and the multivariate regression model selected using cross-validation (SM Table 2).

Table 2: **Results of the univariate and multivariate linear regression of symptoms and activity. The coefficients are the estimated effect on activity when the symptom is present. The multivariate model was selected with a sequential forward floating selection, minimizing the RMSE on test data through a 5-fold cross validation (20 times repeated). 95%CI = The 95% confidence interval for the coefficient.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent: Activity Level |  | Mean (sd) | Coefficient (univariable) | Coefficient (multivariable) |
| Abdominal Pain | No | 4.4 (2.6) | - | - |
|  | Yes | 3.4 (2.7) | -1.01 (-1.90 to -0.12, p=0.026) | - |
| Breathlessness | No | 4.3 (2.7) | - | - |
|  | Yes | 4.1 (2.5) | -0.21 (-0.81 to 0.38, p=0.477) | - |
| Chest Congestion | No | 4.7 (2.9) | - | - |
|  | Yes | 4.0 (2.4) | -0.65 (-1.24 to -0.07, p=0.029) | - |
| Chest Pain | No | 4.4 (2.7) | - | - |
|  | Yes | 4.0 (2.6) | -0.38 (-0.99 to 0.23, p=0.217) | - |
| Chills/Sweats | No | 5.8 (2.8) | - | - |
|  | Yes | 4.1 (2.6) | -1.75 (-2.64 to -0.86, p<0.001) | -0.90 (-1.79 to -0.02, p=0.046) |
| Cough | No | 4.2 (3.2) | - | - |
|  | Yes | 4.3 (2.6) | 0.10 (-1.16 to 1.36, p=0.875) | - |
| Diarrhea | No | 4.4 (2.6) | - | - |
|  | Yes | 3.6 (2.6) | -0.73 (-1.60 to 0.15, p=0.103) | - |
| Ear Pain | No | 4.4 (2.6) | - | - |
|  | Yes | 3.7 (2.7) | -0.69 (-1.44 to 0.05, p=0.068) | - |
| Eye Pain | No | 4.2 (2.6) | - | - |
|  | Yes | 4.4 (2.9) | 0.21 (-0.61 to 1.02, p=0.619) | - |
| Fatigue | No | 5.7 (2.5) | - | - |
|  | Yes | 4.2 (2.6) | -1.53 (-2.72 to -0.34, p=0.012) | - |
| Headache | No | 5.6 (2.8) | - | - |
|  | Yes | 4.0 (2.6) | -1.57 (-2.35 to -0.80, p<0.001) | -1.23 (-1.97 to -0.49, p=0.001) |
| Sleeplessness | No | 4.9 (2.7) | - | - |
|  | Yes | 3.8 (2.5) | -1.14 (-1.71 to -0.57, p<0.001) | -0.91 (-1.44 to -0.37, p=0.001) |
| Itchy Eyes | No | 4.4 (2.7) | - | - |
|  | Yes | 3.7 (2.5) | -0.72 (-1.40 to -0.04, p=0.038) | - |
| Myalgia | No | 5.4 (2.9) | - | - |
|  | Yes | 4.1 (2.6) | -1.32 (-2.25 to -0.38, p=0.006) | - |
| Nasal Congestion | No | 4.4 (2.7) | - | - |
|  | Yes | 4.2 (2.6) | -0.22 (-0.93 to 0.49, p=0.542) | - |
| Nausea | No | 4.7 (2.7) | - | - |
|  | Yes | 3.6 (2.4) | -1.08 (-1.67 to -0.49, p<0.001) | - |
| Sore Throat | No | 4.4 (2.6) | - | - |
|  | Yes | 4.2 (2.7) | -0.20 (-0.95 to 0.55, p=0.605) | - |
| Runny Nose | No | 4.6 (2.6) | - | - |
|  | Yes | 4.1 (2.7) | -0.50 (-1.14 to 0.15, p=0.129) | - |
| Sneeze | No | 4.6 (2.7) | - | - |
|  | Yes | 4.0 (2.6) | -0.67 (-1.24 to -0.09, p=0.024) | - |
| Subjective Fever | No | 5.2 (2.5) | - | - |
|  | Yes | 3.9 (2.6) | -1.25 (-1.90 to -0.60, p<0.001) | -0.67 (-1.32 to -0.01, p=0.047) |
| Swollen Lymph Nodes | No | 4.4 (2.7) | - | - |
|  | Yes | 4.0 (2.5) | -0.46 (-1.05 to 0.13, p=0.128) | - |
| Tooth Pain | No | 4.3 (2.6) | - | - |
|  | Yes | 4.1 (2.7) | -0.24 (-0.98 to 0.50, p=0.526) | - |
| Vomiting | No | 4.5 (2.6) | - | - |
|  | Yes | 2.8 (2.2) | -1.64 (-2.48 to -0.81, p<0.001) | -1.39 (-2.18 to -0.60, p=0.001) |
| Weakness | No | 6.6 (2.4) | - | - |
|  | Yes | 4.1 (2.6) | -2.49 (-3.72 to -1.25, p<0.001) | -1.50 (-2.69 to -0.31, p=0.014) |
| Wheezing | No | 4.4 (2.7) | - | - |
|  | Yes | 4.0 (2.5) | -0.46 (-1.07 to 0.16, p=0.144) | - |

## Distribution of scores

The infectiousness score is skewed left with more than half of the patients having a score of 4 or 5 (SM Figure 2A). The morbidity score is more centered with no patient having a score of 0,1,19 or 20 (SM Figure 2B).

Figure 2: (A) Histogram of infectiousness score for patients with a lab diagnosis of influenza. (B) Histogram of morbidity score for patients with a lab diagnosis of influenza.

Figure 2: (A) Histogram of infectiousness score for patients with a lab diagnosis of influenza. (B) Histogram of morbidity score for patients with a lab diagnosis of influenza.

## Correlation between number and severity of symptoms

The data we have available reports most of the symptoms as absence or presence. If more symptoms correlate positively with stronger symptoms, using absence/presence data only is a valid, albeit less powerful approach. More problematically is a situation in which the number and strength of symptoms correlate negatively. For example, a person having fewer symptoms might “make up” for the lower number of symptoms by having a more severe form. Such a relation would invalidate our approach of adding presence/absence symptoms to arrive at a given score. While we cannot thoroughly test potential correlations between the quantity and strength of symptoms on all our data, we do have severity information on the cough, weakness, and body aches symptoms. This allows us to evaluate the relation between the number of symptoms and the severity of those symptoms.

We find that as the number of infectiousness-related symptoms (i.e., our infectiousness scores) increases, there is an increase in cough intensity (SM Figure 3). The same relationship is found between the number of morbidity-related symptoms (i.e., our morbidity scores) and the severity of weakness and body aches (SM Figures 4 and 5). While not conclusive, this evidence suggests that our assumption that symptom absence/presence also captures symptom severity seems to be defensible.

Figure 3: Relationship between cough intensity and the infectiousness score from main text using all Y/N variables plausibliy related to infectiousness.

Figure 3: Relationship between cough intensity and the infectiousness score from main text using all Y/N variables plausibliy related to infectiousness.

Figure 4: Relationship between weakness intensity and the morbidity score from main text using all Y/N variables related to morbidity.

Figure 4: Relationship between weakness intensity and the morbidity score from main text using all Y/N variables related to morbidity.

Figure 5: Relationship between body ache intensity and the morbidity score from main text using all Y/N variables related to morbidity.

Figure 5: Relationship between body ache intensity and the morbidity score from main text using all Y/N variables related to morbidity.

## Alternative approaches to calculate the infectiousness score

In the main text, we use all of the available yes/no symptoms that can feasibly be related to infectiousness. Since there is no consensus about which symptoms or routes of transmission are the most important, we also consider how robust the results are using several different versions of infectiousness scores.

While chest congestion might be a proxy for pathogen load, i.e., more congestion could indicate higher levels of a pathogen, in which case it could impact infectiousness. However, it is true that chest congestion alone without coughing or sneezing might not lead to increased infectiousness (though some recent studies suggest that breathing alone accounts for a large fraction of expelled influenza virions [1]). To explore if inclusion or exclusion of chest congestion in the infectiousness scores made a difference, we created a score excluding it to see if the overall conclusions would change. The new score values ranged from 0 to 4 (SM Figure 6A).

The same overall relationship is observed between the new infectiousness score and activity. There is still a curve. Spearman’s rank correlation indicates a negative relationship with ( -0.11 (95% CI: -0.22, -0.00)) and the Cochran-Mantel-Haenszel trend test is statistically significant ( 3.772, 1, 0.05) (SM Figure 6B).

When compared to the morbidity score, the same overall trends from the primary analysis were observed. Spearman’s rank correlation ( 0.17 (95% CI: 0.06, 0.27)) and the Cochran-Mantel-Haenszel trend test is statistically significant ( 11.323, 1, < 0.01). We find a mean 1.5 fold increase in the new infectiousness score going from the lowest to the highest morbidity score (SM Figure 6C).

Figure 6: Infectiousness score without chest congestion

Figure 6: Infectiousness score without chest congestion

How patients define congestion symptoms is very subjective [2], so we created a score excluding both of the congestion-related variables to see if the overall conclusions would change. The new score could have a value of 0 to 3 (SM Figure 7A).

The same overall relationship is observed between the new infectiousness score and activity. The curved relationship is not as clear but still present. Spearman’s rank correlation indicates negative relationship with ( -0.11 (95% CI: -0.22, -0.00)) and the Cochran-Mantel-Haenszel trend test is statistically significant ( 4.599, 1, 0.03) (SM Figure 7B).

When compared to the morbidity score again, the same trends were observed, and Spearman’s rank correlation indicates positive relationship ( 0.16 (95% CI: 0.05, 0.26)) and the Cochran-Mantel-Haenszel trend test is statistically significant ( 10.023, 1, < 0.01). We find a mean 1.5 fold increase in the new infectiousness score going from the lowest to the highest morbidity score (SM Figure 7C).

Figure 7: Infectiousness score without chest and nasal congestion variables

Figure 7: Infectiousness score without chest and nasal congestion variables

For the score used in the main text, we included all relevant symptoms even if they can be considered strongly related. This might lead to potential double-counting of some symptoms. To evaluate if this might cause problems, we created an alternative score that removed highly correlated variables based on two cut off values of Yule’s Q [3]. There is no commonly used value to define a cut-off at which correlated variables should be removed. We decided to use an absolute correlation of 0.9 or 0.75 for our cut off. When using the 0.9 cut off only very strongly correlated variables were removed, while for the 0.75 even intermediately strong correlations lead to the removal of one of the variables [4]. For any pair of symptoms with a positive or negative correlation of 0.9 or greater, the more informative symptom (having the proportion Yes/No closest to 50%) was kept[5]. For our data, cough and chest congestion had a Yule’s Q of 0.933, and chest congestion had the best balance between yes and no responses, so it was kept. When relaxing the correlation to 0.75, there is no change in the variables included in the score.

The score values range from 0 to 4 (SM Figure 8A). The same overall relationship is observed between the new infectiousness score and activity. The curve is present, Spearman’s rank correlation indicates negative relationship with ( -0.16 (95% CI: -0.26, -0.05)) and the Cochran-Mantel-Haenszel trend test is statistically significant ( 7.138, 1, < 0.01) (SM Figure 8B).

When compared to the morbidity score again the same overall trends were observed with Spearman’s rank correlation indicates positive relationship ( 0.28 (95% CI: 0.18, 0.38)) and the Cochran-Mantel-Haenszel trend test is statistically significant ( 25.818, 1, < 0.01). We find a mean 2.1 fold increase in the new infectiousness score going from the lowest to the highest morbidity score (SM Figure 8C).

Figure 8: Infectiousness score removing variables determined to be redundant based on Yule’s Q

Figure 8: Infectiousness score removing variables determined to be redundant based on Yule’s Q

## Alternative approaches to calculate morbidity score

To see the impact of removing correlated variables from the morbidity score, we used the same methods applied to the infectiousness score above. Among the morbidity symptoms, only vomiting and weakness correlated greater than 0.9. Vomiting was included in the score since it was more balanced then weakness, which was present in 94% of patients. This score using 0.9 cut off had a possible range of 0 to 19. There are no patients with a morbidity score of 0,1,18,19 (SM Figure 9A). The same relationships observed in the main text are seen when the morbidity score is compared to activity and infectiousness scores.

There is a negative correlation between the morbidity score, and the patient’s self-reported activity level suggests that higher morbidity score is associated with reduced activity levels. Spearman’s rank correlation indicates negative relationship ( -0.31 (95% CI: -0.41, -0.21)) and the Cochran-Mantel-Haenszel trend test is statistically significant ( 36.004, 1, < 0.01). We find a mean 4.3 fold decrease in activity level going from the lowest to the highest morbidity score (SM Figure 9B).

There is a positive correlation between the morbidity and infectiousness scores show a positive correlation. Spearman’s rank correlation indicates positive relationship ( 0.28 (95% CI: 0.18, 0.38)) and the Cochran-Mantel-Haenszel trend test is statistically significant ( 25.942, 1, < 0.01). We find a mean 1.7 fold in the infectiousness score going from the lowest to the highest morbidity score (SM Figure 9C).

Figure 9: Morbidity score removing variables determined to be redundant using the absolute value of Yule’s Q of 0.9 or more

Figure 9: Morbidity score removing variables determined to be redundant using the absolute value of Yule’s Q of 0.9 or more

The score created using the 0.75 cut off was different, with a total of 6 symptoms being excluded. Starting with the highest correlations first: Weakness/Vomit (Q=1) keep vomit, Tooth pain/Headache (Q=.86) keep Tooth pain, Chills Sweats/SoreThroat (Q=-.8) keep Chills Sweats, Fatigue/BodyAches (Q=.80) keep BodyAches, Vomit/Nausea (Q=.79) keep Nausea, SubjectiveFever/ChillsSweats (Q=.76) keep SubjectiveFever. The 0.75 cut off morbidity score included Subjective Fever, Myalgia, Sleeplessness, Breathlessness, Wheezing, Chest pain, Abdominal Pain, Diarrhea, Nausea, Ear Pain, Tooth pain, Eye pain, Itchy Eyes, and Swollen Lymph Nodes. This morbidity score has a possible value of 0 to 14. Compared to the morbidity score used in the main text and 0.9 cut off there are now 12 patients with a score of 0 or 1, but there are still no patients with a score of 13 or 14. (SM Figure 10A). Again the overall results remain the same despite using a different version of the score.

There is a negative correlation between the morbidity score and the patient’s self-reported activity level. Spearman’s rank correlation indicates negative relationship ( -0.26 (95% CI: -0.36, -0.16)) and the Cochran-Mantel-Haenszel trend test is statistically significant ( 24.987, 1, < 0.01). We find a mean 2.9 fold reduction in activity level going from the lowest to the highest morbidity score (SM Figure 10B).

The relationship between the morbidity and infectiousness scores show a positive correlation. Spearman’s rank correlation indicates positive relationship ( 0.29 (95% CI: 0.18, 0.38)) and the Cochran-Mantel-Haenszel trend test is statistically significant ( 25.942, 1, < 0.01). We find a mean 1.7 fold increase in the infectiousness score going from the lowest to the highest morbidity score (SM Figure 10C).

Figure 10: Morbidity score removing variables determined to be redundant using the absolute value of Yule’s Q of 0.75 or more

Figure 10: Morbidity score removing variables determined to be redundant using the absolute value of Yule’s Q of 0.75 or more

# Analysis using patients with any diagnosis influenza (Lab or Empiric)

## Description of the population

Influenza diagnosis for our population is determined using three different methods a rapid antigen test, a PCR test, or an empirical diagnosis by a physician. In the main text, we considered any person who was diagnosed by either a rapid antigen or PCR test as having influenza. Here we repeat the analyses completed in the main text with the addition of patients with a diagnosis of influenza empirically based on symptoms. Patients with an empirical diagnosis are generally defined as having influenza-like illness (ILI). In total, there are 735 patients with any diagnosis of influenza. Among these patients, the activity levels ranged from 0 to 10, with a median of 4. All of the patients had symptoms of disease. The most common symptom is weakness, and the least common symptom is vomiting (SM Table 3).

Table 3: **Out of the 735 patients included the table shows the number of patients who reported having the following symptoms and the corresponding percentage.**

|  |  |
| --- | --- |
|  | Overall |
| n | 735 |
| Abdominal Pain = Yes (%) | 93 (12.7) |
| Breathlessness = Yes (%) | 297 (40.4) |
| Chest Congestion = Yes (%) | 409 (55.6) |
| Chest Pain = Yes (%) | 234 (31.8) |
| Chills/Sweats = Yes (%) | 604 (82.2) |
| Cough = Yes (%) | 660 (89.8) |
| Diarrhea = Yes (%) | 99 (13.5) |
| Ear Pain = Yes (%) | 162 (22.0) |
| Eye Pain = Yes (%) | 113 (15.4) |
| Fatigue = Yes (%) | 671 (91.3) |
| Headache = Yes (%) | 620 (84.4) |
| Itchy Eyes = Yes (%) | 182 (24.8) |
| Myalgia = Yes (%) | 656 (89.3) |
| Nasal Congestion = Yes (%) | 565 (76.9) |
| Nausea = Yes (%) | 258 (35.1) |
| Runny Nose = Yes (%) | 524 (71.3) |
| Sleeplessness = Yes (%) | 419 (57.0) |
| Sneeze = Yes (%) | 395 (53.7) |
| Sore Throat = Yes (%) | 614 (83.5) |
| Subjective Fever = Yes (%) | 505 (68.7) |
| Swollen Lymph Nodes = Yes (%) | 314 (42.7) |
| Tooth Pain = Yes (%) | 166 (22.6) |
| Vomiting = Yes (%) | 79 (10.7) |
| Weakness = Yes (%) | 686 (93.3) |
| Wheezing = Yes (%) | 221 (30.1) |

## Univariate and subset selection

We explored the univariate correlations between activity level and each symptom. All of the symptoms that were statistically significantly related to activity showed a negative correlation with activity level (SM Table 4). Based on the cross-validated variable selection, we found that a model that included chills/sweats, subjective fever, headache, weakness, sleeplessness, and vomiting creates the most predictive model (SM Table 4).

Table 4: **Results of the univariate and multivariate linear regression of symptoms and activity. The coefficients are the estimated effect on activity when the symptom is present. The multivariate model was selected with a sequential forward floating selection, minimizing the root mean square error on test data through a 5-fold cross validation (20 times repeated). 95%CI = The 95% confidence interval for the coefficient.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent: Activity Level |  | Mean (sd) | Coefficient (univariable) | Coefficient (multivariable) |
| Abdominal Pain | No | 4.6 (2.6) | - | - |
|  | Yes | 3.8 (2.7) | -0.78 (-1.35 to -0.20, p=0.008) | - |
| Breathlessness | No | 4.6 (2.7) | - | - |
|  | Yes | 4.2 (2.6) | -0.39 (-0.78 to 0.00, p=0.052) | - |
| Chest Congestion | No | 4.7 (2.7) | - | - |
|  | Yes | 4.2 (2.5) | -0.49 (-0.88 to -0.11, p=0.012) | - |
| Chest Pain | No | 4.6 (2.6) | - | - |
|  | Yes | 4.2 (2.8) | -0.42 (-0.83 to -0.01, p=0.044) | - |
| Chills/Sweats | No | 6.2 (2.5) | - | - |
|  | Yes | 4.1 (2.5) | -2.08 (-2.56 to -1.61, p<0.001) | -1.30 (-1.79 to -0.81, p<0.001) |
| Cough | No | 4.9 (2.8) | - | - |
|  | Yes | 4.4 (2.6) | -0.48 (-1.11 to 0.15, p=0.137) | - |
| Diarrhea | No | 4.6 (2.7) | - | - |
|  | Yes | 3.7 (2.5) | -0.85 (-1.41 to -0.29, p=0.003) | - |
| Ear Pain | No | 4.5 (2.6) | - | - |
|  | Yes | 4.2 (2.7) | -0.34 (-0.80 to 0.12, p=0.149) | - |
| Eye Pain | No | 4.5 (2.7) | - | - |
|  | Yes | 4.5 (2.6) | 0.05 (-0.48 to 0.58, p=0.855) | - |
| Fatigue | No | 5.5 (2.6) | - | - |
|  | Yes | 4.4 (2.6) | -1.15 (-1.83 to -0.48, p=0.001) | - |
| Headache | No | 5.6 (2.6) | - | - |
|  | Yes | 4.3 (2.6) | -1.34 (-1.86 to -0.82, p<0.001) | -0.91 (-1.39 to -0.43, p<0.001) |
| Sleeplessness | No | 5.0 (2.7) | - | - |
|  | Yes | 4.0 (2.5) | -0.96 (-1.35 to -0.58, p<0.001) | -0.72 (-1.07 to -0.37, p<0.001) |
| Itchy Eyes | No | 4.5 (2.7) | - | - |
|  | Yes | 4.4 (2.5) | -0.07 (-0.52 to 0.37, p=0.742) | - |
| Myalgia | No | 5.5 (2.7) | - | - |
|  | Yes | 4.3 (2.6) | -1.14 (-1.75 to -0.53, p<0.001) | - |
| Nasal Congestion | No | 4.8 (2.6) | - | - |
|  | Yes | 4.4 (2.7) | -0.39 (-0.84 to 0.07, p=0.096) | - |
| Nausea | No | 4.8 (2.7) | - | - |
|  | Yes | 3.8 (2.5) | -0.99 (-1.39 to -0.60, p<0.001) | - |
| Sore Throat | No | 4.5 (2.6) | - | - |
|  | Yes | 4.5 (2.6) | -0.02 (-0.54 to 0.50, p=0.939) | - |
| Runny Nose | No | 4.6 (2.7) | - | - |
|  | Yes | 4.4 (2.6) | -0.19 (-0.61 to 0.23, p=0.382) | - |
| Sneeze | No | 4.6 (2.7) | - | - |
|  | Yes | 4.3 (2.6) | -0.27 (-0.66 to 0.11, p=0.164) | - |
| Subjective Fever | No | 5.6 (2.5) | - | - |
|  | Yes | 4.0 (2.6) | -1.62 (-2.01 to -1.22, p<0.001) | -0.92 (-1.33 to -0.51, p<0.001) |
| Swollen Lymph Nodes | No | 4.5 (2.6) | - | - |
|  | Yes | 4.4 (2.7) | -0.13 (-0.52 to 0.25, p=0.495) | - |
| Tooth Pain | No | 4.6 (2.6) | - | - |
|  | Yes | 4.2 (2.7) | -0.40 (-0.86 to 0.05, p=0.084) | - |
| Vomiting | No | 4.6 (2.6) | - | - |
|  | Yes | 3.1 (2.3) | -1.57 (-2.18 to -0.96, p<0.001) | -1.27 (-1.84 to -0.71, p<0.001) |
| Weakness | No | 6.3 (2.5) | - | - |
|  | Yes | 4.3 (2.6) | -1.98 (-2.73 to -1.22, p<0.001) | -0.92 (-1.64 to -0.21, p=0.011) |
| Wheezing | No | 4.7 (2.7) | - | - |
|  | Yes | 4.0 (2.5) | -0.67 (-1.09 to -0.26, p=0.001) | - |

## Computation of infectiousness and morbidity scores

We used the same symptom classification presented in the main text. The median infectiousness score is 4, and the a similar skewed distribution is present with most of the patients having a score of 4 or 5 (SM Figure 11).

Figure 11: Distribution of the infectiousness score.

Figure 11: Distribution of the infectiousness score.

The median morbidity score is 9, and no patients have a morbidity score of 0, 1, 19, 20 (SM Figure 12). Such a centered distribution is assumed to be a result of patients felling ill enough to seek medical care, but none were sick enough to require urgent care or hospitalization.

Figure 12: Distribution of the morbidity score.

Figure 12: Distribution of the morbidity score.

## Impact of infectiousness score on activity

Analysis of the impact of the infectiousness score on activity suggests that the value of this score has a negative correlation with the activity level. Spearman’s rank correlation is -0.10 (95% CI: -0.17, -0.03) and the Cochran-Mantel-Haenszel trend test is statistically significant ( 7.083, 1, < 0.01) (SM Figure 13). This is different from the main analysis were we did not observe a clear relationship between activity and the infectiousness score.

Figure 13: Activity level for each level of the infectiousness score. The red diamond is the mean. The solid blue line is the linear regression fit. The shaded area is the 95% confidence interval for the linear regression.

Figure 13: Activity level for each level of the infectiousness score. The red diamond is the mean. The solid blue line is the linear regression fit. The shaded area is the 95% confidence interval for the linear regression.

## Impact of morbidity score on activity

Analysis of the impact of the morbidity score on activity suggests that the value of this score is correlated with the activity level of a patient, with higher morbidity correlating with reduced activity. Spearman’s rank correlation indicates a negative relationship -0.32 (95% CI: -0.38, -0.25) and the Cochran-Mantel-Haenszel trend test is statistically significant ( 78.501, 1, < 0.01) (SM Figure 14). The observed pattern is clear, with a mean 3.6 fold decrease in activity level going from the lowest to the highest morbidity score.

Figure 14: Activity level for each level of the morbidity score. The red diamond is the mean. The solid blue line is the linear regression fit. The shaded area is the 95% confidence interval for the linear regression.

Figure 14: Activity level for each level of the morbidity score. The red diamond is the mean. The solid blue line is the linear regression fit. The shaded area is the 95% confidence interval for the linear regression.

## Impact of morbidity score on infectiousness score

Analysis of the relationship between the morbidity and infectiousness scores show a positive correlation. Spearman’s rank correlation indicates a positive relationship ( 0.26 (95% CI: 0.19, 0.33)) and the Cochran-Mantel-Haenszel trend test is statistically significant ( 44.505, 1, < 0.01) (SM Figure 15). Apart from the values activity levels for low morbidity score (with small sample sizes), the pattern is consistent with a mean 1.7 fold increase in the infectiousness score going from the lowest to the highest morbidity score.

Figure 15: Infectiousness score for each level of the morbidity score. The red diamond is the mean. The solid blue line is the linear regression fit. The shaded area is the 95% confidence interval for the linear regression.

Figure 15: Infectiousness score for each level of the morbidity score. The red diamond is the mean. The solid blue line is the linear regression fit. The shaded area is the 95% confidence interval for the linear regression.

# References

1. Yan J *et al.* 2018 Infectious virus in exhaled breath of symptomatic seasonal influenza cases from a college community. *Proceedings of the National Academy of Sciences* **115**, 1081–1086.

2. McCoul ED, Mohammed AE, Debbaneh PM, Carratola M, Patel AS. 2019 Differences in the Intended Meaning of Congestion Between Patients and Clinicians. *JAMA Otolaryngology–Head & Neck Surgery* **145**, 634–640. (doi:[10.1001/jamaoto.2019.1023](https://doi.org/10.1001/jamaoto.2019.1023))

3. Yule GU. 1919 *An introduction to the theory of statistics*. C. Griffin, limited.

4. Warrens MJ. 2008 On association coefficients for 2 2 tables and properties that do not depend on the marginal distributions. *Psychometrika* **73**, 777.

5. Kuhn M, others. 2008 Building predictive models in r using the caret package. *Journal of Statistical Software* **28**, 1–26.