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

2019-08-18

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, Department of Epidemiology and Biostatistics, The University of Georgia, Athens, GA, 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)

# Documentation for additional SM Files

## Getting the Files

Files described below are in the Supplemental Materials zip file.

* “**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.
* “**2 Data Cleaning Script**” folder contains two R scripts that merge and clean 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 included but will not run since none of the raw data sets are 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 Univariate 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 Univariate Plots.R**” This script creates all of the plots. Results are saved in “5 Results”
  + “**Flu Symptoms Activity Univariate 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)
  + “**american-journal-of-epidemiology.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)
  + “**american-journal-of-epidemiology.csl**” is a style file to format the citations in the supplement

## Reproducing Results

The files required to reproduce the results consists of 2 R Markdown files, 4 R script, and one anonymized data file. These files allow 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 supplementary zip file 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 three clean data sets used for all further analyses. The data sets are all saved in the “3 Clean Data” folder and include:
   1. “SympAct” Contains data for all the patients with a respiratory primary complaint.
   2. “SympAct\_Any\_Pos.Rda” Contains data for all influenza patients regardless of diagnosis method.
   3. “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\_Any\_Pos.Rda” and “SympAct\_Lab\_Pos.Rda” are both subsets of “SympAct.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 Material for Manuscript

##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.

##Correlation of symptoms reported in the main text

### Infectiousness symptom correlation

Cough and chest congestion had a Yule correlation coefficient greater than 0.9(SM Figure 2).

Figure 2: Correlation of infectiousness symptoms for patients with a lab diagnosis of influenza.

Figure 2: Correlation of infectiousness symptoms for patients with a lab diagnosis of influenza.

### Morbidity symptom correlation

Vomiting and weakness had a Yule correlation coefficient greater than 0.9 (SM Figure 3).

Figure 3: Correlation of morbidity symptoms for patients with a lab diagnosis of influenza.

Figure 3: Correlation of morbidity symptoms for patients with a lab diagnosis of influenza.

# Sensitivity Analyses

## Correlation Cut off of 0.75 vs. 0.9

### Summary of differences

The overall conclusions and the infectiousness score did not change at all. The morbidity score changed with 7 symptoms being excluded. This new morbidity score included Abdominal Pain, Breathlessness, Chest pain, Diarrhea, Ear Pain, Headache, Itchy Eyes, Myalgia, Nausea, Sleeplessness, Subjective Fever, Swollen Lymph Nodes, and Wheezing. The new morbidity score had a possible range of 0 to 13.

The distribution is similar in that is centered but there is a difference in the minimum and maximum score (1 compared to 2 and 11 compared to 17 respectively) (SM Figure 4). The observed relationship the morbidity score had between activity and infectiousness score are unchanged (SM Figure 5 and 6).

### Calculating new morbidity score

The morbidity score did change. When the cut off of 0.75 was applied, seven symptoms were dropped. Starting with the highest correlations first: Weakness/Vomit (Q=1) keep vomit, Tooth pain/Headache (Q=.87) keep Headache, Headache/Eye pain (Q=.83) keep Headache, swollen lymph nodes/SoreThroat (Q=.81) keep SwollenLympnodes, Fatigue/Myalgia (Q=.80) keep BodyAches, SubjectiveFever/ChillsSweats (Q=.78) keep SubjectiveFever, Vomit/Nausea (Q=.77) keep Nausea. The new morbidity score includes Abdominal Pain, Breathlessness, Chest pain, Diarrhea, Ear Pain, Headache, Itchy Eyes, Myalgia, Nausea, Sleeplessness, Subjective Fever, Swollen Lymph Nodes, and Wheezing. The new morbidity score ranges from 0 to 13.

The mean morbidity score when 0.75 was used as the cut off was 5.51, and no patients had a morbidity score of 0, 12, or 13 (SM Figure 4). The distribution is still as expected since all the patients felt ill enough to seek medical care, but none were sick enough to require urgent care or hospitalization.

Figure 4: Distribution of the morbidity score.

Figure 4: Distribution of the morbidity score.

###Impact of morbidity score on activity Analysis of the association between the new morbidity score and the patient’s self-reported activity level suggests that higher morbidity score is associated with a reduced activity. Spearman’s rank correlation indicates a negative relationship -0.33 (95% CI: -0.42, -0.23) and the Cochran-Mantel-Haenszel trend test is statistically significant ( 36.78, 1, < 0.01) (SM Figure 5). The observed pattern is consistent and clear, with a reduction of 67% in mean activity going from the lowest to the highest morbidity score.

Figure 5: Activity level for each level of the morbidity score. Red diamonds indicate the mean. The solid blue line is the linear regression fit. The shaded area is the 95% confidence interval for the linear regression.

Figure 5: Activity level for each level of the morbidity score. Red diamonds indicate 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 0.28 (95% CI: 0.18, 0.38) and the Cochran-Mantel-Haenszel trend test is statistically significant ( 25.52, 1, < 0.01) (SM Figure 6). Apart from the activity levels for low morbidity score values (with small sample sizes), the pattern is consistent and clear, with an increase of 33% in the infectiousness score going from the lowest to the highest morbidity score.

Figure 6: Infectiousness score for each level of the morbidity score. Red diamonds indicate the mean. The solid blue line is the linear regression fit. The shaded area is the 95% confidence interval for the linear regression.

Figure 6: Infectiousness score for each level of the morbidity score. Red diamonds indicate the mean. The solid blue line is the linear regression fit. The shaded area is the 95% confidence interval for the linear regression.

## Analysis using all patients diagnosed with influenza

### Summary of differences

The overall conclusions remain the same when the empirically diagnosis patients are included. From here on the population used to generate the results in the main text will be referenced to as “lab diagnosis” and the population to generate the results below will be referenced to as “any diagnosis”

There were no meaningful differences in the univariate analysis (SM Table 2). Among patients with any diagnosis the most predictive multi-variate model was different then the model selected using lab diagnosis and included chest congestion, headache, sleeplessness, subjective fever, vomiting, and weakness (SM Table 2). Both models included 6 symptoms, and 5 of the symptoms are in both (headache, sleeplessness, subjective fever, vomiting, and weakness). For patients with any diagnosis chills/sweats was included while chest congestion was included for the patients with a lab diagnosis.

Both of the scores were different for the any diagnosis patients compared to the lab diagnosis patients. The infectiousness score for the any diagnosis patients included all of the possible symptoms (SM Figure 7, 9), compared to the lab diagnosed patients were cough was removed. Among the morbidity symptoms for patients with any diagnosis none were had a correlation greater than 0.9. Compared to the patients with a lab diagnosis were the morbidity score excluded weakness. Based on these results two new scores were calculated for the patients with any diagnosis. The infectiousness score had a possible range of 0 to 5, and the morbidity score had a possible range of 0 to 20.

Using the new scores we examined the relationships of the scores between each other and activity levels. We again found that the infectiousness score had a weak association with reported activity, while the morbidity score showed a clear correlation with both the reported activity level and infectiousness score (SM Figures 11 - 13 ).

### Description of the population

Influenza diagnosis for our population is determined using three different methods; a rapid antigen test, a PCR test, or by a physician giving an empirical diagnosis. 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 716 patients with any diagnosis of influenza. These Patients reported activity levels ranging from 0 to 10 with a mean of 4.46. All of the patients had symptoms of disease with only 16% reporting 10 or fewer. The most common symptom is weakness, and the least common symptom is vomiting (SM Table 1).

Table 1: **Out of the 716 patients included the table shows the number of patients who reported having the following symptoms and the corresponding percentage.**

|  |  |
| --- | --- |
|  | Overall |
| n | 716 |
| Abdominal Pain = Yes (%) | 91 (12.7) |
| Breathlessness = Yes (%) | 287 (40.1) |
| Chest Congestion = Yes (%) | 398 (55.6) |
| Chest Pain = Yes (%) | 224 (31.3) |
| Chills/Sweats = Yes (%) | 589 (82.3) |
| Cough = Yes (%) | 646 (90.2) |
| Diarrhea = Yes (%) | 98 (13.7) |
| Ear Pain = Yes (%) | 158 (22.1) |
| Eye Pain = Yes (%) | 112 (15.6) |
| Fatigue = Yes (%) | 653 (91.2) |
| Headache = Yes (%) | 604 (84.4) |
| Itchy Eyes = Yes (%) | 179 (25.0) |
| Myalgia = Yes (%) | 637 (89.0) |
| Nasal Congestion = Yes (%) | 550 (76.8) |
| Nausea = Yes (%) | 254 (35.5) |
| Runny Nose = Yes (%) | 511 (71.4) |
| Sleeplessness = Yes (%) | 409 (57.1) |
| Sneeze = Yes (%) | 388 (54.2) |
| Sore Throat = Yes (%) | 598 (83.5) |
| Subjective Fever = Yes (%) | 493 (68.9) |
| Swollen Lymph Nodes = Yes (%) | 308 (43.0) |
| Tooth Pain = Yes (%) | 163 (22.8) |
| Vomiting = Yes (%) | 79 (11.0) |
| Weakness = Yes (%) | 667 (93.2) |
| Wheezing = Yes (%) | 217 (30.3) |

### 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 2). Based on 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 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 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) | -0.79 (-1.37 to -0.21, p=0.008) | - |
| Breathlessness | No | 4.6 (2.7) | - | - |
|  | Yes | 4.2 (2.6) | -0.37 (-0.77 to 0.02, p=0.066) | - |
| Chest Congestion | No | 4.7 (2.7) | - | - |
|  | Yes | 4.2 (2.5) | -0.49 (-0.88 to -0.10, p=0.013) | - |
| Chest Pain | No | 4.6 (2.6) | - | - |
|  | Yes | 4.1 (2.8) | -0.45 (-0.87 to -0.03, p=0.035) | - |
| Chills/Sweats | No | 6.2 (2.6) | - | - |
|  | Yes | 4.1 (2.5) | -2.07 (-2.55 to -1.58, p<0.001) | -1.27 (-1.77 to -0.77, p<0.001) |
| Cough | No | 4.8 (2.8) | - | - |
|  | Yes | 4.4 (2.6) | -0.43 (-1.08 to 0.22, p=0.196) | - |
| Diarrhea | No | 4.6 (2.7) | - | - |
|  | Yes | 3.7 (2.5) | -0.82 (-1.38 to -0.26, p=0.004) | - |
| Ear Pain | No | 4.5 (2.6) | - | - |
|  | Yes | 4.2 (2.6) | -0.35 (-0.82 to 0.12, p=0.143) | - |
| Eye Pain | No | 4.4 (2.6) | - | - |
|  | Yes | 4.5 (2.6) | 0.04 (-0.49 to 0.58, p=0.876) | - |
| Fatigue | No | 5.5 (2.6) | - | - |
|  | Yes | 4.4 (2.6) | -1.19 (-1.87 to -0.51, p=0.001) | - |
| Headache | No | 5.6 (2.6) | - | - |
|  | Yes | 4.2 (2.6) | -1.31 (-1.84 to -0.79, p<0.001) | -0.89 (-1.38 to -0.40, p<0.001) |
| Sleeplessness | No | 5.0 (2.7) | - | - |
|  | Yes | 4.1 (2.5) | -0.94 (-1.32 to -0.55, p<0.001) | -0.68 (-1.04 to -0.32, p<0.001) |
| Itchy Eyes | No | 4.5 (2.7) | - | - |
|  | Yes | 4.4 (2.5) | -0.05 (-0.50 to 0.40, p=0.832) | - |
| Myalgia | No | 5.5 (2.7) | - | - |
|  | Yes | 4.3 (2.6) | -1.15 (-1.77 to -0.54, p<0.001) | - |
| Nasal Congestion | No | 4.8 (2.6) | - | - |
|  | Yes | 4.4 (2.7) | -0.39 (-0.85 to 0.07, p=0.098) | - |
| Nausea | No | 4.8 (2.7) | - | - |
|  | Yes | 3.8 (2.5) | -0.97 (-1.37 to -0.58, p<0.001) | - |
| Sore Throat | No | 4.5 (2.7) | - | - |
|  | Yes | 4.4 (2.6) | -0.07 (-0.60 to 0.45, p=0.782) | - |
| Runny Nose | No | 4.6 (2.7) | - | - |
|  | Yes | 4.4 (2.6) | -0.15 (-0.58 to 0.27, p=0.479) | - |
| Sneeze | No | 4.6 (2.7) | - | - |
|  | Yes | 4.4 (2.6) | -0.22 (-0.61 to 0.17, p=0.273) | - |
| Subjective Fever | No | 5.6 (2.5) | - | - |
|  | Yes | 3.9 (2.6) | -1.64 (-2.04 to -1.24, p<0.001) | -0.94 (-1.35 to -0.53, p<0.001) |
| Swollen Lymph Nodes | No | 4.5 (2.6) | - | - |
|  | Yes | 4.4 (2.6) | -0.09 (-0.48 to 0.30, p=0.643) | - |
| Tooth Pain | No | 4.5 (2.6) | - | - |
|  | Yes | 4.2 (2.7) | -0.34 (-0.81 to 0.12, p=0.145) | - |
| Vomiting | No | 4.6 (2.6) | - | - |
|  | Yes | 3.1 (2.3) | -1.56 (-2.17 to -0.96, p<0.001) | -1.27 (-1.83 to -0.71, p<0.001) |
| Weakness | No | 6.3 (2.5) | - | - |
|  | Yes | 4.3 (2.6) | -1.99 (-2.74 to -1.23, p<0.001) | -0.94 (-1.66 to -0.22, p=0.010) |
| Wheezing | No | 4.7 (2.7) | - | - |
|  | Yes | 4.0 (2.5) | -0.69 (-1.11 to -0.27, p=0.001) | - |

### Computation of Transmission and Morbidity Scores

We used the same symptom classification presented in the main text.

None of the symptoms related to infectiousness were correlated with each other at a level of greater than 0.9 (SM Figure 7). This result differs from that in the main text where were cough was excluded. A new infectiousness score was calculated for this population ranging from 0 to 5.

Among the morbidity symptoms none had a correlation greater than 0.9 (SM Figure 8). This result differs from the analysis in the main text where vomiting was retained, and weakness was excluded. A new morbidity score was calculated for this population ranging from 0 to 20.

Figure 7: Correlation of infectiousness symptoms for patients with any diagnosis of influenza

Figure 7: Correlation of infectiousness symptoms for patients with any diagnosis of influenza

Figure 8: Correlation of morbidity symptoms for patients with any diagnosis of influenza

Figure 8: Correlation of morbidity symptoms for patients with any diagnosis of influenza

The median infectiousness score is 4, and only 13 patients have an infectiousness score of 0 (SM Figure 9). Only 23% of patients have a score of 3 or less (SM Figure 9).

Figure 9: Distribution of the infectiousness score.

Figure 9: 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 10). Such a centered distribution is expected since all the patients felt ill enough to seek medical care, but none were sick enough to require urgent care or hospitalization.

Figure 10: Distribution of the morbidity score.

Figure 10: 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 -0.09 (95% CI: -0.17, -0.02) and the Cochran-Mantel-Haenszel trend test is statistically significant ( 5.94, 1, 0.01) (SM Figure 11). This is different from the main analysis were we did not observe a clear relationship between activity and the infectiousness score.

Figure 11: 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 11: 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 ( 76.04, 1, < 0.01) (SM Figure 12). There is a reduction of 80% in mean activity going from the lowest to the highest morbidity score.

Figure 12: 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 12: 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, 0.32)) and the Cochran-Mantel-Haenszel trend test is statistically significant ( 41.66, 1, < 0.01) (SM Figure 13). Apart from the values activity levels for low morbidity score (with small sample sizes), the pattern is consistent and clear, with an increase of 67% in the infectiousness score going from the lowest to the highest morbidity score.

Figure 13: 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 13: 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.