Prediction of Exacerbations in Asthma Using Real World Data (RWD)

by Iris

Abstract

Background: Asthma affects more than 25 million Americans (1). It is characterized by inflammation of the airways in the lung. A person diagnosed with asthma may experience an acute worsening of their condition, called an **exacerbation**, which can be triggered by many reasons including pollen, dust, chemicals, environmental changes or stress, among others. During an exacerbation, the person experiences shortness of breath, cough, wheezing or chest tightness. During course of asthma care, health care providers perform a series of tests to assess how well lungs are functioning (2). These tests measure the "**obstruction**" in the lung i.e. how much and how quickly air is being inhaled or exhaled. This measurement is performed using an instrumental technique called Spirometry (3).

Objective: To better understand triggers of exacerbation, I have analyzed a longitudinal spirometry dataset from patients diagnosed with asthma within a large multi-site healthcare system and have tried to **determine (i) if the level of obstruction in the lung can predict the likelihood of future exacerbations and (ii) if there were any additional, potential confounding factors which need to be considered.** This information will help better understand the bridge between obstruction and exacerbation, predict future exacerbation events and can be used to inform future interventions.

Methods: The data for this study came from a large multi-site healthcare system. In addition to demographics and Body Mass Index (BMI), the data consisted of 2 Spirometry measurements: (i) Forced Vital Capacity (FVC) or the largest amount of air that can forcefully be exhaled after breathing as deeply as possible and (ii) Forced expiratory volume (FEV_1) or the amount of air which can be exhaled in one second.

De-identified data from a cohort of asthma patients with visits from 2009-2013 were collected and formatted for analysis. All data cleaning and analysis was performed using R v. 3.4.3 and RStudio. Demographic and baseline measurement data (sex, BMI, age, FEV_1/FVC) were summarized by exacerbation status. A crude odds ratio for the likelihood of an exacerbation was determined using the FEV_1/FVC ratio alone and the results were compared and analyzed against the adjusted odds ratio for the likelihood of an exacerbation with consideration of other confounding factors. The results were modelled using logistic regression and were further analyzed using Directed Acyclic Graph (DAG) model (4).

Results: **Using logistic regression and DAG model, the odds ratios indicated that the level of obstruction (FEV_1/FVC ratio) alone is not an accurate predictor for an exacerbation.** Through further experimentation, it was discovered that the baseline characteristics of patients, such as **age, gender, and BMI act as important confounders and are useful in predicting an exacerbation based on obstruction**. These results were identified using the logistic regression and DAG models for both crude and adjusted effects of obstruction on likelihood of exacerbation.

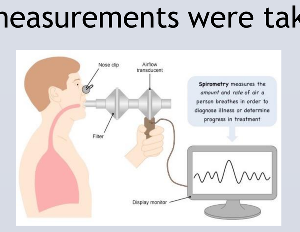
Conclusion: The results obtained from this experiment are useful in creating a more accurate predictive model for the likelihood of an exacerbation. Eventually, these findings can aid in the development of a formula which can calculate a new categorizable number for spirometry ratios with the consideration of the confounding factors of age, gender, and BMI. A shortcoming of this experiment is the limited demographics of the real-world data as the data set is taken from a generally "obstructed" population.

Introduction

1. Real-World-Data (RWD)

In the Health Industry, Real-World-Data (RWD) typically refers to data collected from multiple sources on patients, outside of clinical trials. The sources of these data could be social media, Electronic Medical or Health Records forms (EMR/EHR forms), personal devices, insurance claims, pharmacy prescriptions, health networks and many others. These data may then be used by others to analyze and obtain critical insights including deeper understanding of the diseases, treatment compliance, or other, lesser-known effects of a drug etc. etc.

In this project, I used RWD for patients diagnosed with asthma who visited a particular healthcare facility for their regular check-ups and follow-ups. Each time a patient went to the doctor in this facility, a spirometry test was done and 2 measurements were taken:



(i) Forced Vital Capacity (FVC) or the largest amount of air that can forcefully be exhaled after breathing as deeply as possible

(ii) Forced expiratory volume (FEV_1) or the amount of air which can be exhaled in one second

Studies have shown that measured FEV_1 vs FVC are valuable indicators of whether or not the said patient is "obstructed" or not. A FEV_1/FVC ratio ≥0.7 is considered normal

	Normal	Mild Obstruction	Moderate Obstruction	Severe Obstruction
FEV1/FVC	≥ 0.7	0.6 0 – 0.69	0.50 - 0.59	< 0.5

2. Predictive Analytics and DAG Model

Most literature describes Predictive Analytics as "a branch of analytics which is used to make predictions about unknown future events". Predictive analytics uses methods like data mining, statistics, modeling, machine learning and artificial intelligence to analyze current data to make predictions about the future (predictiveanalyticstoday.com). These methods are used by scientists in many facets of life e.g. prediction of weather changes, consumer behavior, stock market etc.

In this project,

- I used obstruction status of asthmatic patients (normal, mild, moderate or severe; represented by FEV_1/FVC) to predict the likelihood of future exacerbation
- I identified a few confounding factors which can influence and improve the ability of FEV_1/FVC to predict the likelihood of future exacerbation
- I used crude and adjusted odds ratios to demonstrate the effectiveness of these newly identified confounding factors in predicting the likelihood of future exacerbation based on obstruction (FEV_1/FVC)

Methods

Data Collection: The data for this study came from a large multi-site integrated delivery network (IDN) in Utah in de-identified format. Its electronic medical record (EMR) data contained, in addition to spirometry data, information from outpatient clinics (primary and specialty care) and hospitals including clinical visits, hospitalizations, outpatient prescriptions, inpatient pharmacy data, and lab data. The observation window for the study was the period from 01/01/2009 to 12/31/2013 in which patients must have been observed for at least one year.

Analysis: The following steps were used to clean and format the data, identify potential confounders and compute crude and adjusted odds ratios for exacerbation. Free source online tools like R v.3.4.3 scripts (https://cran.r-project.org/bin/windows/base/), Rstudio (https://www.rstudio.com) and DAG (http://www.dagitty.net/) were downloaded and used for the project

Step 1: Truncate and convert data into readable form for Rstudio

Step 2: (i) Calculate mean±SD for FEV_1/FVC and other variables (age, gender, and BMI) based on exacerbation status (ii) Represent data in graphs to show distribution → Table 1, Figure 1 & Figure 2

Step 3: Calculate crude odds ratio for likelihood of exacerbation based on FEV_1/FVC, using logistic regression → Table 2 (Column #1)

Step 4: Calculate adjusted odds ratio by incorporating confounding factors identified via DAGs, i.e., age, gender, and BMI → Figure 3 & Table 2 (Column #2)

Step 5: Compare the results from the un-adjusted and adjusted logistic models → Figure 4 & Figure 5

Results

Table 1: Demographic Characteristics of Patients by Exacerbation Status

Variable	No Exacerbation	Exacerbation
	n (%)	n (%)
Ethnicity		
White	1,889 (73.3%)	689 (26.7%)
Black	13 (68.4%)	6 (31.6%)
Other	82 (67.2%)	40 (32.8%)
Gender		
Male	898 (79.8%)	228 (20.2%)
Female	1,086 (68.2%)	507 (31.8%)
	Mean (SD)	Mean (SD)
Age (in yrs)	64.28 (14.54)	62.97 (15.52)
BMI	34.44 (11.88)	38.56 (14.41)
FEV/FVC ratio	0.64 (0.15)	0.65 (0.15)

- Table 1 shows the baseline characteristics of the analysis cohort by exacerbation outcome

Figure 1: Distribution of Obstruction (FEV_1/FVC)

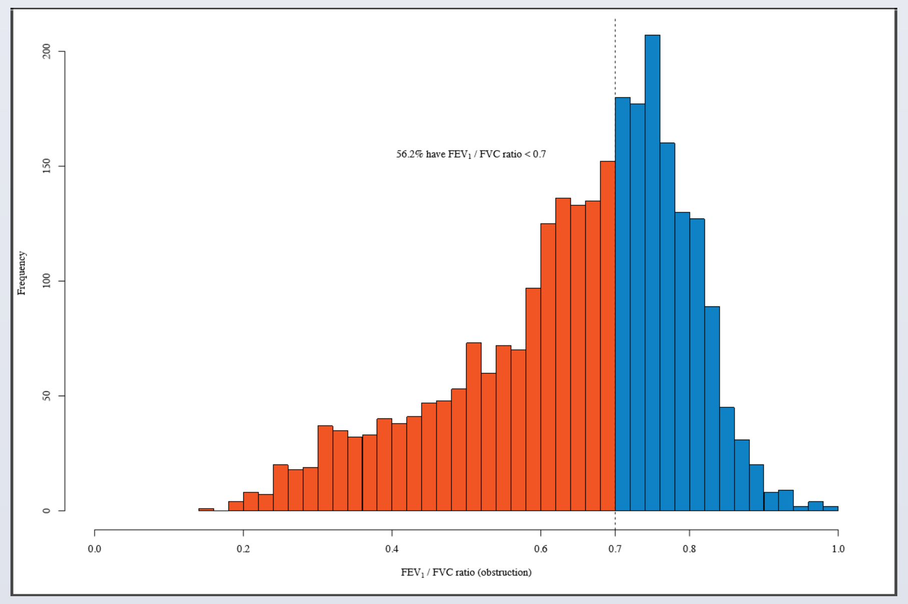


Figure 1 Illustrates the distribution of subjects in the analytic dataset, plotted using their obstruction status (FEV_1/FVC)
- The subjects with FEV_1/FVC < 0.7 are represented in red (0.7 is a researched value for the baseline of obstruction)
- The values in the 'obstructed" cohort show a greater range than non-obstructed (blue)

Figure 2: Scatterplot of FEV_1 vs FVC values

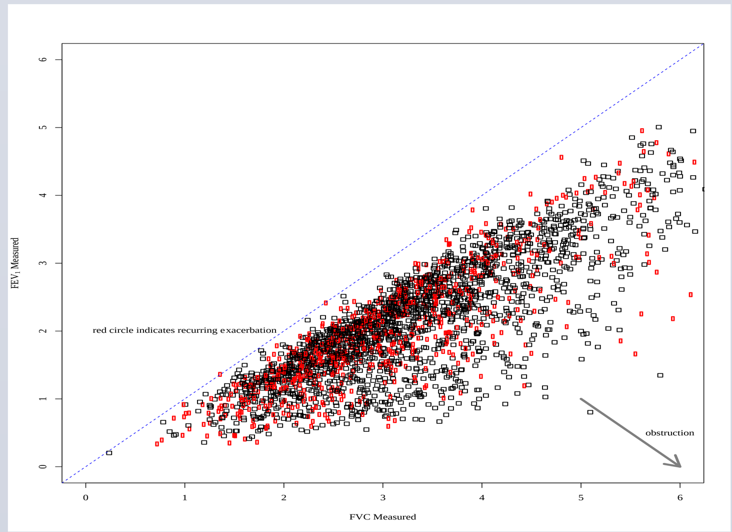
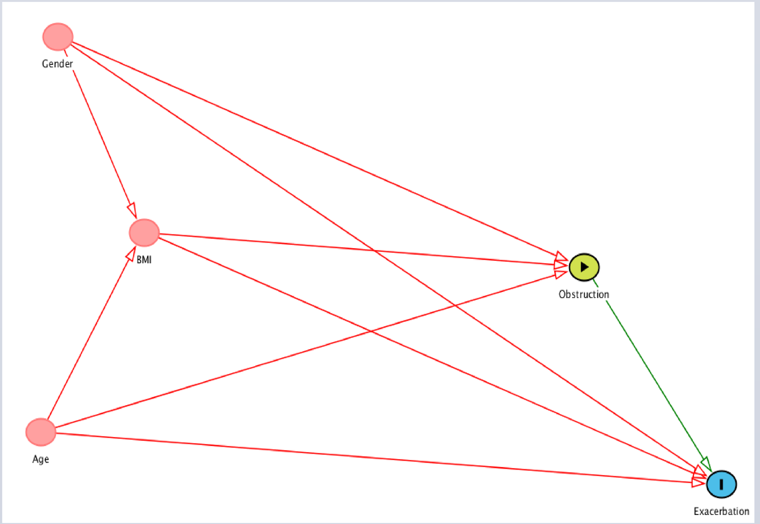


Figure 2 shows the relationship between the measured FEV_1 and FVC values for each patient in the cohort
- red circles indicate exacerbation
- although it is difficult to distinguish signal from noise, there is a trend towards lower FEV_1/FVC ratios

Figure 3: DAG Showing Relationship Between Obstruction and Exacerbation and Potential Confounding Pathways



The DAG indicates that in order to estimate the un-confounded effect of obstruction on exacerbation, an adjustment needs to be made for Age, Gender, and BMI

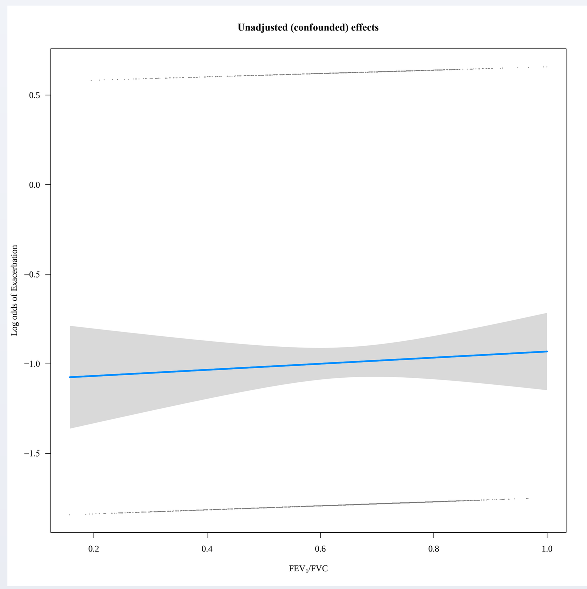
Table 2: Crude and Adjusted Effects of FEV_1/FVC on Exacerbation

Variable	P-value (unadjusted effect)	P-value (adjusted effect)
FEV1/FVC	0.552	0.008
Age	/	0.023
BMI	/	<0.001
Gender	/	<0.001

- Column #2 shows that FEV_1/FVC alone is not a reliable predictor of exacerbation (*p>0.05*)
- Column #3 shows strong correlation between FEV_1/FVC and exacerbation, considering confounding factors (*p<0.05*)

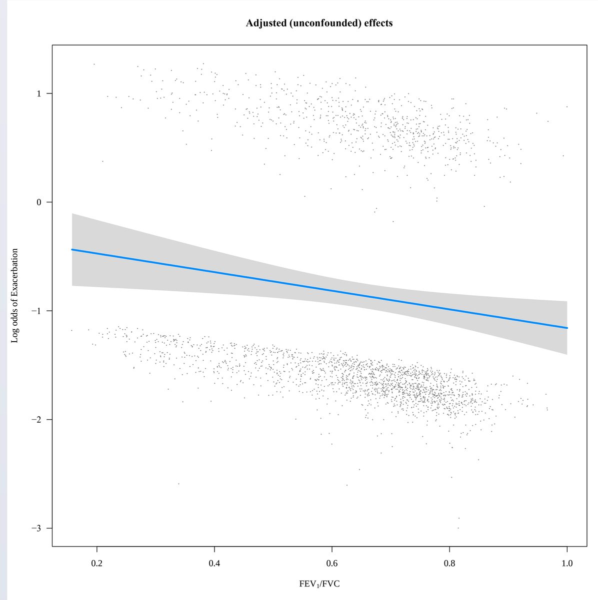
Results (Cont.)

Figure 4: Unadjusted Effect of FEV_1/FVC Ratio on the Odds of Exacerbation



- The un-adjusted effect of obstruction on exacerbation was not significant due to the fairly level logistic regression model
- Crude odds ratio: 1.19 [95% Confidence Interval (CI): 0.68-2.08]
- p-value = 0.552

Figure 5: Adjusted Effect of FEV_1/FVC Ratio on the Odds of Exacerbation

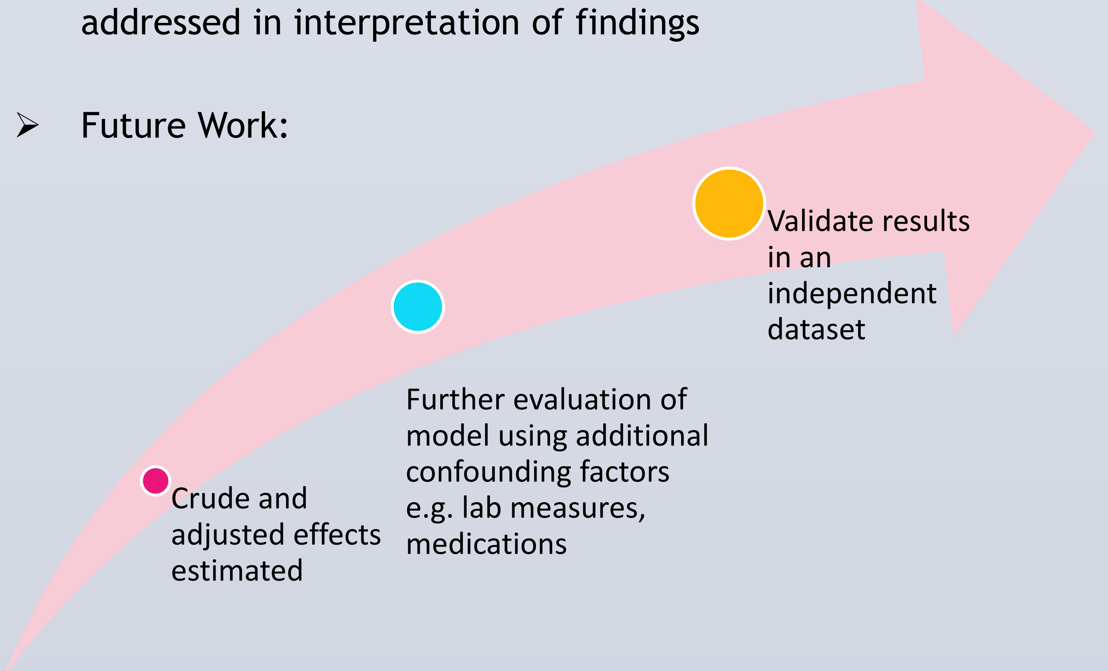


- The adjusted effect of obstruction on exacerbation was statistically significant, as evident from the steeper logistic regression model
- Adjusted odds ratio: 0.42 (95% [95% Confidence Interval (CI): 0.23-0.80]
- p-value = 0.00779

- *P value <0.05 ⇒ "significant" effect*

Conclusion and Future Work

- Obstruction alone was not a significant indicator of exacerbation in this cohort
  - This is likely due to causal pathways, as were later identifiable via DAG
- Several other measures e.g. smoking status, lab results (IgE, eosinophils etc.) were not considered to limit the scope of the project
- This is a medically complex cohort (initial obstruction was high) and likelihood for selection bias exists and should be addressed in interpretation of findings
- Future Work:



References

1. AAAAI. *Asthma.* (2018). Retrieved from http://www.aaaai.org/conditions-and-treatments/asthma
2. Fahy JV. *Airway Mucus and the Mucociliary System*. In: Middleton E Jr, Reed CE, Ellis EF, Adkinson NF Jr, Yunginger JW, Busse WW, editors. Allergy Principles & Practice. 5th ed. St. Louis: Mosby-Year Book, Inc; 1998. pp. 520–531.
3. Mayo Clinic Staff. *Spirometry*. (2017, August 17). Retrieved from https://www.mayoclinic.org/tests-procedures/spirometry/about/pac-20385201
4. Shrier I, Platt RW. *Reducing bias through directed acyclic graphs*. BMC Med Res Methodol. 2008;8:70