|  |
| --- |
| **Graphical User Interface (GUI) Platform to Estimate Chronic Obstructive Pulmonary Disorder (COPD) from User-Given Risk Factors[[1]](#footnote-2)**  Ben Barrett, Jason Huang, Gurpinderjit Judge  1 School of Biomedical Engineering, Drexel University, USA  Course: Bmes550 Advanced Biocomputational Languages  Instructor: Ahmet Sacan  Date: 2022-12-07 |

[[2]](#footnote-3)\*abstract

Chronic Obstructive Pulmonary Disorder (COPD) affects millions of people in the US, with an estimated prevalence of up to 10% of the US population [1]. Several risk factors for developing COPD include age, smoking habits, genetic mutations or the presence of variants, and exercise habits. Literature reviews demonstrate that there are few comprehensive risk calculators, specifically for COPD and designed for both user and physician interaction. Here, we demonstrate the feasibility of a Graphical User Interface (GUI) based platform to deliver a risk score to patients, and an interactive platform for physicians to determine a patient’s changing risk score. Despite the feasibility of the platform itself, future alterations will be needed to validate in the output results for both patients and physicians.

# introduction

COPD has demonstrated to create a significant burden on public health affecting up to 10% of the US population. COPD typically involves the buildup of mucus in airways that prevents lungs from properly functioning [1]. Mucus build-up leads to restricted airflow, causing alveoli collapse, further impairing proper lung functioning, affecting daily activities and even leading to early death. Numerous risk factors exist for COPD development, with smoking habits historically being the leading indicator [1,2]. Recent studies, however, have started to demonstrate the impact that non-smoking factors can have on COPD development [3].

Currently, there are limited means of patients understanding their risk for developing COPD without a detailed discussion with their physician. Physicians can even have trouble understanding the true picture of COPD due to the multitude of possible symptoms and causes. While there are options available to understand risk factors that are more user-friendly, these are generally free, online calculators that have not been peer-reviewed and are therefore, lacking in credibility.

Therefore, we set out to develop a tool that can be used simultaneously by both physicians and patients that can be used as a guide to COPD risk factors. Patients will be able to input values for their current risk factors including smoking history, family history, age, exercise habits and receive a score output. Physicians, however, will have much more involved capabilities due to their training in using patient data. Once physicians input a password, they will be given a chart depicting how a patient’s risk will change over time if their behaviors remain unchanged but will also be able to input changes to certain factors and understand how this affects a patient’s risk of developing COPD. Such a feature would allow a physician to have a data driven discussion with the patient to help them convey how behavioral changes will have significant impacts on their health. Since this data is subject to misinterpretation by patients, they will be unable to view this feature.

# Dataset

Data based on conclusions from various studies will be incorporated into the development of a SQL database for recovery by the calculator program. In Eun-Jung Kim et al., smoking status was evaluated through a lifestyle intervention program to measure the occurrence of COPD. The Korea National Health and Nutrition Examination Survey, administered from 2005 to 2015, analyzed the effect of aging and cigarette smoking on COPD. It was concluded that the two risk factors have a strong positive correlation with COPD (Eun-Jung Kim et al., 2019). In another study, the predictive capacity of family history and polygenic risk score on COPD and COPD-related outcomes was analyzed on non-Hispanic white and African American subjects. The data was drawn from COPDGene and ECLIPSE studies. It was found that a family history of COPD, defined as a maternal or paternal history of COPD, had a positive correlation with COPD risk, but not as strong, however (Moll et al., 2020). In a study performed by Yan et al., from 2005-2009, 46,285 individuals between the ages of 35-70 were recruited from urban and rural communities in China were surveyed to access their physical activity intensity as well as other factors on diagnosed COPD patients. It was determined that proper physical exercise can prevent COPD (Yan et. al., 2017). Lastly, genetic variant data from cases of bronchus and lung diseases was pulled from the National Cancer Institute and the top three gene variants, being “TP53,” “TTN,” and “MUC16” were utilized (GDC Data). As such, all this information we gathered was implemented into the risk calculator for COPD. A database schema is included below in Figure 1.

Diagram

Description automatically generated

**Figure 1. Database ER Schema**. Only two tables were used to connect the data in this study, one for behavioral risk factors and one for the Genetic Mutations. The data was developed to simulate relative risk between factors and isolated using SQL Select statements.

# methods and IMplementation

To develop our program, we implemented a series of GUI apps from the MATLAB App Designer and MATLAB functions to perform SQL queries and statistical analysis. The purpose of the software is two-fold, to allow a patient to understand their current risk of COPD, and for the physician to use the software as a guide for diagnosis.

What must first be discussed is the SQL based processing of data to determine a proper multiple linear regression. As there is limited data on comprehensive prevalence, physicians would theoretically be able to upload data and determine inferred risk from this dataset. A database will be constructed and then processed using a combination of SQL queries and table processing functions within MATLAB to use the   
“regress” function within MATLAB to develop a formula that will be used to determine the risk scores and graphs that will be displayed to the users.

We will be using a common GUI page that will be used to direct the user to their corresponding role of either patient or physician and allow them to enter the necessary inputs. There will be two separate GUI files used for each role since each requires different inputs to give different outputs.

The workflow of the calculator diverges from this point. If the user is a patient, they will be directed to patientrisk.mlapp and be able to input their risk factors into the GUI, which will feed these inputs into a multiple linear regression calculating MATLAB script, Riskcalculator.m. The Riskcalculator.m function will return a score for the patientrisk.mlapp to display.

Alternatively, the patient may declare themselves to be a physician, the user will be asked to input a password. This feature is intended to prevent people from attempting to use the physician pathway without the necessary credentials. Once a password has been entered and accepted, the user will be taken to a separate GUI file, physicianrisk.mlapp. For the time being, this password is hard-coded and cannot be changed by the user. When the physician reaches the physicianrisk.mlapp GUI, they will be able to input a patient’s risk factors, with the exception of their file being that they can select from the previously discussed variants, to see a plot of the patient’s associated risk over time and the current comprehensive risk score.

Once the plot and score panels appear, a fourth panel will appear allowing the physician to alter three of the risk factors: smoking habits, physical activity and genetic variant. Genetic variant is included with behavioral risk factors only for the possibility that targeted genetic therapies emerge allowing the risk of this variant to change. The behavioral risk factors are included so that the physician can offer advice to the patients about the controllable factors leading to COPD risk.

As for the calculations, a separate script was used to calculate the multiple regression and those values were hard coded into the risk calculator function. We recognize this is a limitation because new data cannot be uploaded, since it is not needed for the out purposes.

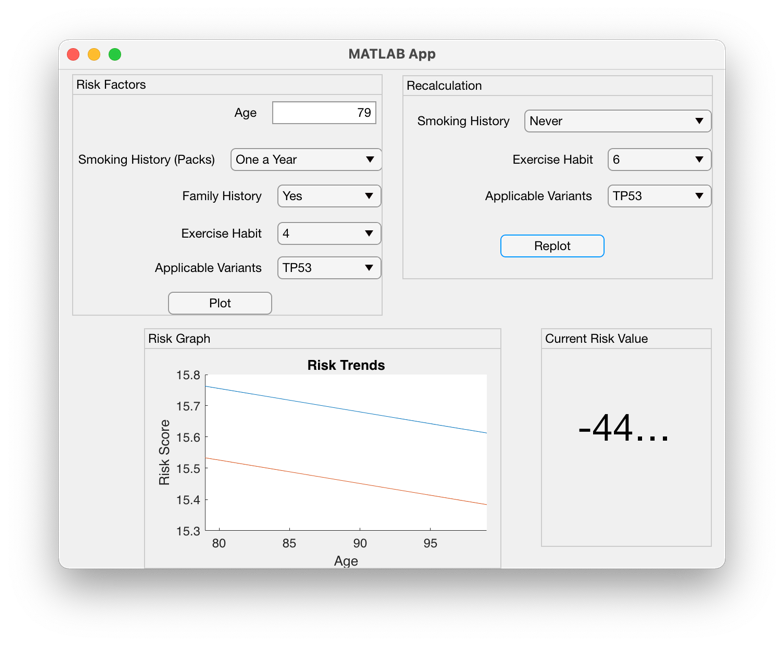
# Experiments and REsults

To determine the efficacy of the model for our program, we used a test case on a patient that is 79 Years Old, smokes one pack of cigarettes per week, exercises an average of 2 days per week, and is positive for the TP53 and does have family history of COPD. The data will be input into the physician interface as well as the patient interface to determine if the program correctly models the changes in risk score with altered behavior.

As can be seen in Figure 1., the initial set of data from the patient presents a graph and an initial score. Currently, the trend demonstrates that the patient’s risk is modeled to decrease with age, which is not incredibly consistent with previous research on COPD. Figure 2. demonstrates the effect of changing certain behavioral risk factors, by decreasing smoking habits and increasing the number of days of exercise. As expected, the second (orange) line represents a decreased risk of COPD over time.

Graphical user interface

Description automatically generated



**Figure 2. Initial Display of Physician Interface.** Illustrates what a physician will be displayed once they enter the initial set of data. Currently, the trend illustrated is inconsistent with expectations since the risk with age should increase.

The same patient’s data was input into the patient portal and a risk score was produced as demonstrated in Figure 3. and compared to a patient of the same age, but with less frequent smoking and more exercise for comparison if trends are biologically accurate.

# Graphical user interface Description automatically generated

**Figure 3. Second Display of Physician Interface.** Two key risk factors have been altered, Smoking History and Exercise Habit to illustrate a consistently decreased score in the patient’s future. The plot from altered behavioral factors is demonstrated by the orange line.

Expectedly, the results in Figure 3. are biologically representative, which indicates that the calculation methods may only need subtle revision. As age increases, and improved factors are added, the risk score decreases.

We also ran a scenario where a similar aged patient, with family history had less smoking and more exercise days, which should theoretically improve their risk score. Figure 4., shows that this was successful, even if by a small margin, demonstrating proof of concept.

**Graphical user interface

Description automatically generatedGraphical user interface

Description automatically generated**

**Figure 4. Patient Portal Comparing Patient #1 and Patient #2.** A Risk Score is calculated for the Patients with the Demographics discussed above. Expectedly, the Risk Value for Patient #2 with higher

# DISCUSSION

The calculator is successful in incorporating inputs from GUI files into a MATLAB script that will produce

a risk score for patients and trends in risk as patients age. The calculator also demonstrates that a user with the physician role can alter the factors of a patient and examine how these alterations affect a patient’s risk over time. The trends seen in the data displayed, do make biological sense based on a literature review conducted prior to the implementation of the calculator. As risk factors associated with a higher prevalence of disease are included, the associated risk score increases within the function. Conversely, as the risk factors are improved, the patient’s risk of developing COPD decreases.

As the results demonstrate, there are several limitations that future iterations and studies could address to improve the calculator’s function. First, correcting the trends seen in the data to ensure that there is biological sense. The efficacy of the value provided for the calculator remain up for debate. study’s main limitation is the precision of the data involved. Clinicians and researchers continue to examine what constitutes as a risk factor for COPD, as there is still debate on the exact criteria for COPD. While prevalence was included in determining the associated risk score for each factor, the data is far from comprehensive and requires additional review to perfect an exact “score”.

Additionally, there are several improvements that could made to address some of the software shortcomings in this initial iteration. Currently, the password used for the physicianrisk.mlapp directory is universal and hardcoded, which should be altered in future iterations to allow the physician user to set their own, unique password. The data is also not comprehensive and future iterations should make sure that additional data can be uploaded and used to perfect the coefficients used in the calculation.

However, the features within the program prove promising as they allow common use between both physicians and patients for a similarly common purpose. The potential for this tool is rather unlimited as this can help in a variety of diagnostic setting for physicians but helping them develop their own understanding of the disease, make data driven decisions, and have more transparent communications with their patients. Patients will also get an accurate glimpse into how their actions can have a direct and immediate impact on their health.

# References

Kim,

1. *COPD: Causes, symptoms, diagnosis, treatment & prevention*. Cleveland Clinic. (n.d.). Retrieved December 7, 2022, from https://my.clevelandclinic.org/health/diseases/8709-chronic-obstructive-pulmonary-disease-copd

2. Miravitlles, M., & Ribera, A. (2017). Understanding the impact of symptoms on the burden of COPD. *Respiratory Research*, *18*(1). https://doi.org/10.1186/s12931-017-0548-3

3. BERNICE H. COHEN, WILMOT C. BALL, JR, SHIRLEY BRASHEARS, EARL L. DIAMOND, PAUL KREISS, DAVID A. LEVY, HAROLD A. MENKES, SOLBERT PERMUTT, MELVYN S. TOCKMAN, RISK FACTORS IN CHRONIC OBSTRUCTIVE PULMONARY DISEASE (COPD), *American Journal of Epidemiology*, Volume 105, Issue 3, March 1977, Pages 223–232, <https://doi.org/10.1093/oxfordjournals.aje.a112378>

., 4. Yoon, S.-J., Kim, Y.-E., Go, D.-S., & Jung, Y. (2019). Effects of Aging and Smoking Duration on Cigarette Smoke-Induced COPD Severity. *Journal of Korean Medical Science*, *34*(Suppl 1). https://doi.org/10.3346/jkms.2019.34.e90

1. Moll, M., Lutz, S. M., Ghosh, A. J., Sakorn

sakolpat, P., Hersh, C. P., Beaty, T. H., Dudbridge, F., Tobin, M. D., Mittleman, M. A., Silverman, E. K., Hobbs, B. D., & Cho, M. H. (2020). Relative contributions of family history and a polygenic risk score on COPD and related outcomes: COPDGene and ECLIPSE studies. *BMJ Open Respiratory Research*, *7*(1), e000755. https://doi.org/10.1136/bmjresp-2020-000755

1. Wang, Y., Bo, J., & Li, W. (2017). Healthy lifestyle behaviors among individuals with chronic obstructive pulmonary disease in urban and rural communities in China: a large community-based epidemiological study. *International Journal of Chronic Obstructive Pulmonary Disease*, *Volume 12*(1), 3311–3321. <https://doi.org/10.2147/copd.s144978>
2. GDC Data

https://portal.gdc.cancer.gov/exploration?cases\_offset=80&filters=%7B%22content%22%3A%5B%7B%22content%22%3A%7B%22field%22%3A%22cases.primary\_site%22%2C%22value%22%3A%5B%22bronchus%20and%20lung%22%5D%7D%2C%22op%22%3A%22in%22%7D%2C%7B%22content%22%3A%7B%22field%22%3A%22cases.diagnoses.tissue\_or\_organ\_of\_origin%22%2C%22value%22%3A%5B%22lower%20lobe%2C%20lung%22%2C%22lung%2C%20nos%22%2C%22main%20bronchus%22%2C%22middle%20lobe%2C%20lung%22%2C%22overlapping%20lesion%20of%20lung%22%2C%22upper%20lobe%2C%20lung%22%5D%7D%2C%22op%22%3A%22in%22%7D%5D%2C%22op%22%3A%22and%22%7D&searchTableTab=genes

1. . [↑](#footnote-ref-2)
2. [↑](#footnote-ref-3)