PharmaSUG 2022 - Paper RW-179

GENERATING REAL WORLD EVIDENCE ON THE LIKELIHOOD OF METASTATIC CANCER IN PATIENTS THROUGH MACHINE LEARNING IN OBSERVATIONAL RESEARCH: INSIGHTS FOR PREVENTION

Sherrine Eid, MPH, Samiul Haque, PhD, S. Robert Collins, SAS Institute, Inc.

ABSTRACT

OBJECTIVES Properly identifying comorbidities in cancer patients that increase the likelihood of metastatic disease is critical to preventing disease progression and morbidity. Living with comorbidities can lead to an increased likelihood of negative outcomes in patients who are diagnosed with cancer. This research aims to identify opportunities to prevent or at least affect the likelihood of metastasis.

METHODS Over 2.3 million deidentified, cancer patient medical claims records (www.compile.com) were analyzed to assess the likelihood of metastatic cancer. Cohorts were defined as any patient who was diagnosed with an ICD-10 diagnosis of C7.x, C78.x, C79.x, or C80.x (n=2,396,043). The outcome variable was metastatic cancer defined by Elixhauser. Lasso logistic regression, decision trees, gradient boosting and random forest were run and compared adjusting for sex, age and the top six Elixhauser comorbidity groups (alcohol abuse, congestive heart failure, coagulopathy, anemia, hypertension, and liver disease.) Analyses were executed using SAS® Health.

RESULTS Liver Disease, Hypertension and Coagulopathy, respectively showed a 8.3% to 2.6% magnitude in contributing to the likelihood metastases. Decision tree, random forest, and neural network models (KS Youden 0.0793, 0.009 and 0, respectively) were the least fit models, while the gradient boosting model and logistic regression were the best models. (KS Youden 0.112, and 0.167, respectively)

CONCLUSIONS Electronic medical records should identify patients who have a comorbidity of liver disease to more closely monitor them for metastatic disease. Intervention is crucial to improving metastasis and mortality in these patients.

INTRODUCTION

Properly identifying comorbidities in cancer patients that increase the likelihood of metastatic disease is critical to preventing disease progression and morbidity. Living with comorbidities can lead to an increased likelihood of negative outcomes in patients who are diagnosed with cancer. This research aims to identify opportunities to prevent or at least affect the likelihood of metastasis.

The presence of comorbidity affects the care of cancer patients, many of whom are living with multiple comorbidities. Comorbidity refers to the existence of a long-term health condition in the presence of a primary disease of interest. Having one or more comorbidities may influence the patient's prognosis for a primary disease such as cancer. An association between comorbidity (not specific to any primary disease of interest) and chronic disease outcomes has been reported previously.

METHODS

Over 2.3 million deidentified, cancer patient medical claims records (www.compile.com) were analyzed to assess the likelihood of metastatic cancer. Cohorts were defined as any patient who was diagnosed with an ICD-10 diagnosis of C7.x, C78.x, C79.x, or C80.x (n=2,396,043) in 2017 with a baseline of 12 months and a follow up of 48 months. The outcome variable was metastatic cancer defined by Elixhauser. Lasso logistic regression, decision trees, gradient boosting and random forest were run and compared adjusting for sex, age and the top six Elixhauser comorbidity groups (alcohol abuse, congestive heart failure, coagulopathy, anemia, hypertension, and liver disease.) The conditions, selected following a systematic search of the data, included conditions of the Elixhauser and any highly prevalent conditions that may influence cancer management alone or in combination with another condition.

The International Statistical Classification of Diseases and Related Health Conditions tenth edition (ICD-10) codes captured within the diagnostic fields of the patient records provided information on health

conditions recorded during inpatient and outpatient episodes of care. We used the ICD-10 code groupings of health conditions suggested by Elixhauser to define comorbidities using administrative data.

DESCRIPTIVE DATA ANALYSIS

Descriptive statistics of demographics were included using the Cohort Characterization templated analytic in SAS® Health. Univariate analyses of age, gender, and race were reported.

Analyses were executed using SAS® Health.

RESULTS

Liver disease, hypertension and coagulopathy, respectively, showed a 8.3% to 2.6% magnitude in contributing to the likelihood metastases. Decision tree, random forest, and neural network models (KS Youden 0.0793, 0.009 and 0, respectively) were the least fit models, while the gradient boosting model and logistic regression were the best models. (KS Youden 0.112, and 0.167, respectively)

Error! Reference source not found. Figure 1 shows network analysis performed to examine general relationships among the variables. It revealed a strong relationship between uncomplicated hypertension and solid tumors without metastasis. There were moderate relationships between solid tumors without metastasis and chronic pulmonary disease, obesity, uncomplicated diabetes, and hypothyroidism. Directly related to this research question, metastatic cancer was more strongly related to solid tumor without Metastasis.

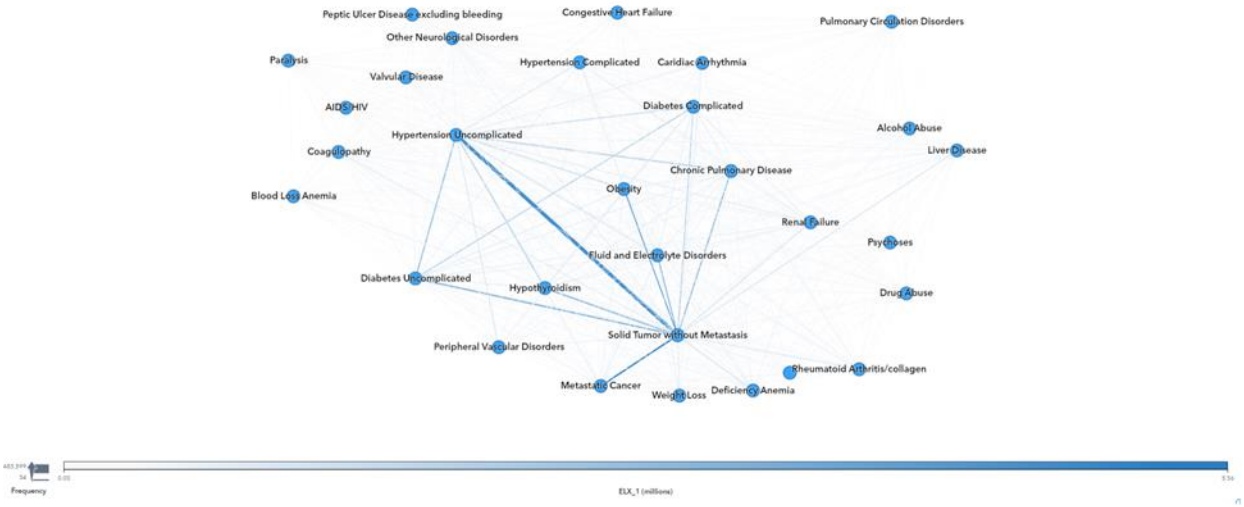

Figure 1 Network Analysis of Elixhauser Comorbidities in Patients Diagnosed with Cancer

Logistic regression analysis demonstrated coagulopathy, liver disease, and pulmonary disorders were the three strongest contributors to the likelihood of metastatic cancer. The model was an adequate fit with a misclassification rate of 7.8% using nearly 2.4 million observations. These analyses were repeated using a gradient boosting model, decision tree, neural network and a forest model.

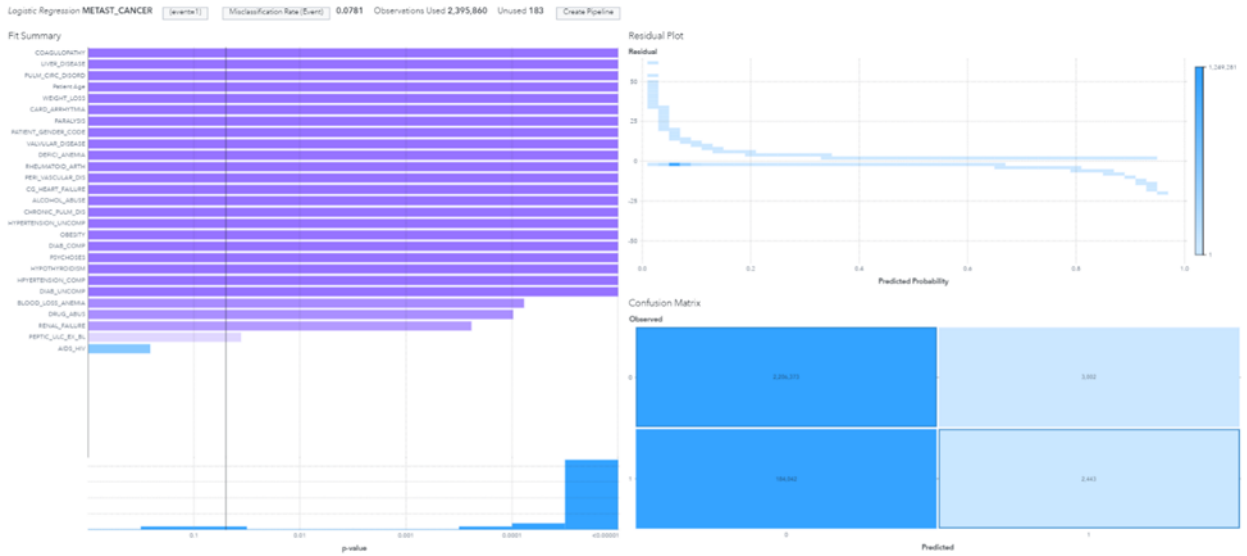

Figure 2 Likelihood of Metastatic Cancer in Patients Diagnosed with Cancer and Other Comorbidities using Logistic Regression.

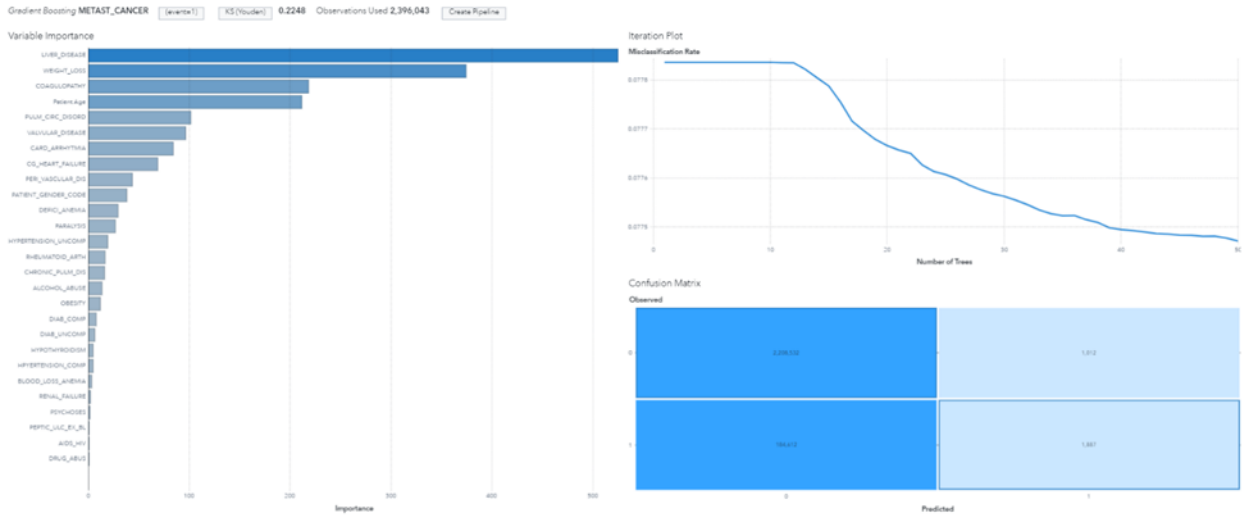

Figure 3 Likelihood of Metastatic Cancer in Patients Diagnosed with Cancer and Other Comorbidities using Gradient Boosting.

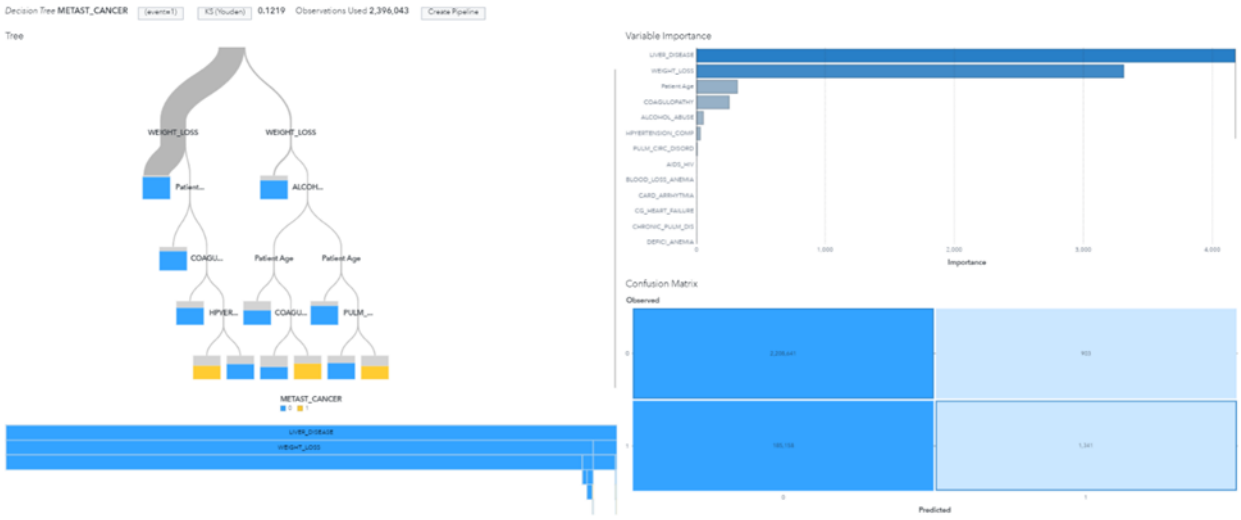

Figure 4 Likelihood of Metastatic Cancer in Patients Diagnosed with Cancer and Other Comorbidities using Decision Tree.

Figure 5 Likelihood of Metastatic Cancer in Patients Diagnosed with Cancer and Other Comorbidities using Forest.

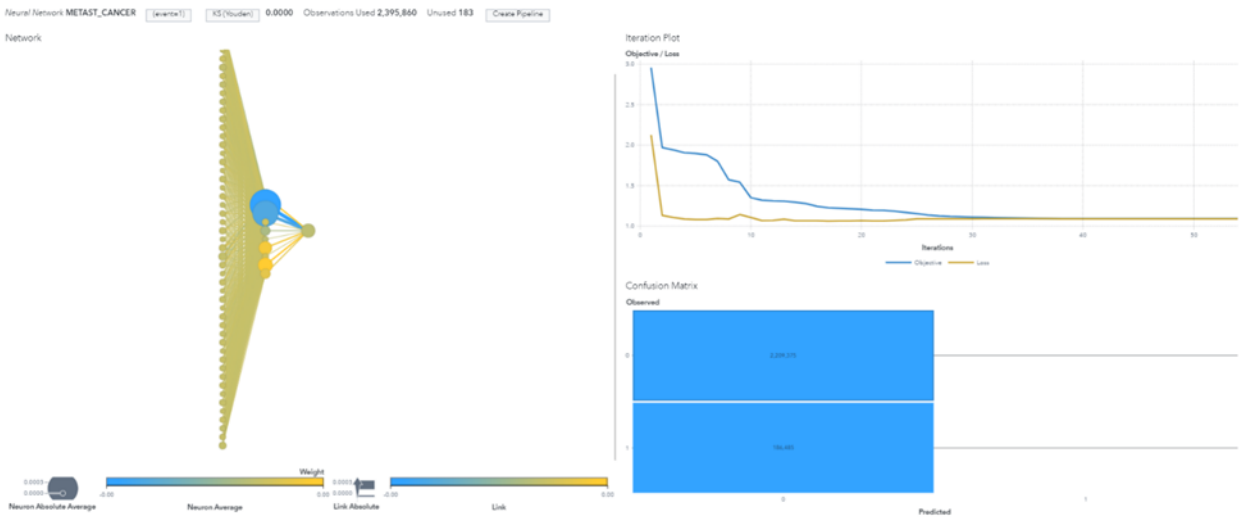

Figure 6 Likelihood of Metastatic Cancer in Patients Diagnosed with Cancer and Other Comorbidities using Neural Networks.

Table 1 shows variables of importance in these models varied from model to model. All models ranked liver disease and weight loss as the two most important variables in the model with the exception of the Logistic Regression model which listed coagulopathy and liver disease the two strongest contributors to the likelihood of metastatic cancer.

Variable Importance	Logistic Regression	Gradient Boosting	Decision Tree	Forest
1	Coagulopathy	Liver Disease	Liver Disease	Liver Disease
2	Liver Disease	Weight Loss	Weight Loss	Weight Loss
3	Pulmonary Circulatory Disorder	Coagulopathy	Age	Coagulopathy
4	Age	Age	Coagulopathy	Pulmonary Circulatory Disorder
5	Weight Loss	Pulmonary Circulatory Disorder	Alcohol Abuse	Age

Table 1. Variable of Importance Order by Model

Models were compared to identify the best fit model for the data using YS (Youden) to select the best fit model (see Figure 7). Gradient boosting model was the best fit model for these data.

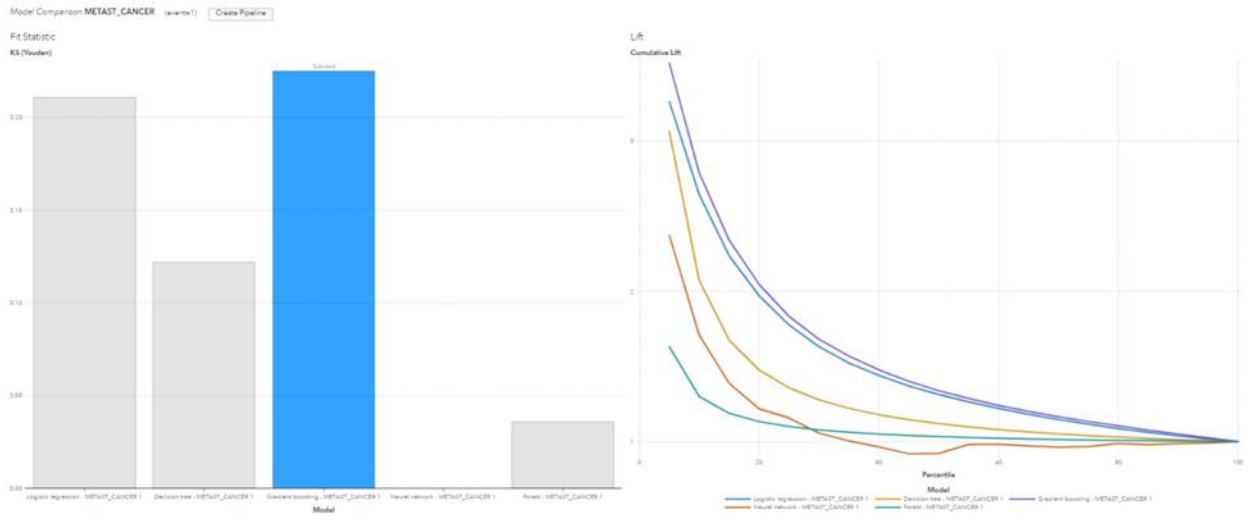

Figure 7 Model Comparison and Selection using KS (Youden).

CONCLUSION

Today, patients create more data than ever that is effectively captured. These data are often stored as "digital exhaust" in the patients' electronic medical or health records (EMR/EHR), medical and pharmacy claims transactions, labs, registries and smart devices (phones, tablets, or other personal devices). Leveraging these Real World Data in observational research allows healthcare and medical researchers to come a step closer to understanding the "whole patient" and generating Real World Evidence.

Our study examined various dimensions of comorbidities as defined by an establish metric, Elixhauser Comorbidity Risk Score, and their potential role in contributing to the likelihood of a patient diagnosed with cancer developing metastatic cancer. As discussed in the results, newer and established methods in predictive analytics and machine learning were leveraged to examine these relationships. Observational research and analytical methods are continuously evolving to address the limitations of these data. Although logistic regression is very commonly used in medical research, we demonstrated that a newer method in machine learning such as gradient boosting is a better fit model for these data.

As healthcare research continues to adapt to more data and different types, formats and sources of data methods and researchers must adapt as well. Leveraging best in class analytics and novel applications will yield better discoveries upon which healthcare professionals can take action to improve and save the lives of their patients and the public.

Our study suggests that there is a significant role that liver disease plays in the likelihood of metastatic cancer. Future activities could include developing a machine learning algorithm where a patient's risk score is calculated and surfaced in their EMR/EHR at the point of care to assist their healthcare provider in determining the best care plan for them.

Electronic medical records should identify patients who have a comorbidity of liver disease to more closely monitor them for treatment response and disease progression to metastatic disease. Intervention is crucial to improving metastasis and mortality in these patients.

REFERENCES

Lüftner D, Hartkopf AD, Lux MP, et al. Challenges and Opportunities for Real-World Evidence in Metastatic Luminal Breast Cancer. Breast Care (Basel). 2021;16(2):108-114. doi:10.1159/000515701

Pérol D, Robain M, Arveux P, Mathoulin-Pélissier S, Chamorey E, Asselain B, et al. The ongoing French metastatic breast cancer (MBC) cohort: the example-based methodology of the Epidemiological Strategy and Medical Economics (ESME) BMJ Open. 2019 Feb;9((2)):e023568.

Jansana A, Del Cura I, Prados-Torres A The SURBCAN group, et alUse of real-world data to study health services utilisation and comorbidities in long-term breast cancer survivors (the SURBCAN study): study protocol for a longitudinal population-based cohort studyBMJ Open 2020;10:e040253. doi: 10.1136/bmjopen-2020-040253

Ewertz M, Land LH, Dalton SO, et al. Influence of specific comorbidities on survival after early-stage breast cancer. Acta Oncol 2018;57:129–34.doi:10.1080/0284186X.2017.1407496

Salisbury C, Johnson L, Purdy S, et al. Epidemiology and impact of multimorbidity in primary care: a retrospective cohort study. Br J Gen Pract 2011;61:e12–21.doi:10.3399/bjgp11X548929

ACKNOWLEDGMENTS

The authors of this paper would like to acknowledge Natallia Drobau for her continued efforts and support in loading the data in the SAS® Health solution.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Sherrine Eid, MPH
SAS Institute,Inc
Sherrine.Eid@sas.com
www.sas.com