



The 65th ASH Annual Meeting Abstracts

ONLINE PUBLICATION ONLY

905.OUTCOMES RESEARCH-LYMPHOID MALIGNANCIES

Utilizing a Machine Learning Approach to Define Racial Difference in Multi-Systemic Impact of Monoclonal Protein in Patient with Multiple MyelomaEhsan Malek, MD¹, Gi-Ming Wang², Jennifer Cullen¹, Curtis Tatsuoka³, Anant Madabhushi¹, James J. Driscoll II, MD PhD⁴¹Case Western Reserve University, Cleveland, OH²Case Western Reserve University, Cleveland³University of Pittsburg, Pittsburg⁴University Hospitals Cleveland Medical Center, Cleveland, OH

Multiple myeloma (MM) is the second common hematologic malignancy characterized by the clonal proliferation of plasma cells. While the clinical outcomes of MM have been improved significantly, there remains an intriguing observation regarding the potential influence of race on disease biology and treatment response. In MM, the malignant plasma cell clone not only disrupts the local bone marrow microenvironment but also exerts a systemic impact on various physiological processes result in significant derangements in immune function, hematopoiesis, renal function, bone metabolism, and electrolyte homeostasis. Understanding the extent of this systemic impact and its potential variation between racial groups is crucial for optimizing patient care and treatment strategies. Here, we *hypothesized* that the systemic impact of the plasma cell clone does not differ significantly between African American and Caucasian patients. We employed a machine learning approach to capture the full spectrum of systemic effects comprehensively. Our novel model was trained to predict the magnitude of the plasma cell clone, as represented by serum monoclonal protein levels. By systematically comparing model performance with and without the inclusion of race as a variable, we sought to evaluate the potential influence of race on the systemic impact of the plasma cell clone.

Method: A total of 171 patients with plasma cell dyscrasias including 1,472 observations were analyzed. where the upper limit of the observed M-spike was 3.5 gr/dL. Forty three clinical and lab variables as predictors of M-spike were fed into the machine learning model. Two lagged variables as the last two preceding M-spike values by the same subject were included. The random forest model was used, where regression forests are an ensemble of different regression trees and are used for nonlinear multiple regression. The goal of using a large number of trees was to train enough that each feature had a chance to appear in several models. The data was randomly split into a training set (80%) and a test set (20%), and a regression tree was built with the training set and then validated using the test set. Bootstrapping was used to generate a collection of data sets (n=500). Importance was measured by leaving a covariate out of models and comparing performance with its inclusion.

Result: The training set used to develop the ML algorithm comprised 749 patient-based observations, which represented ~50% of the patient-based observations. The number of observations at each M-Spike value and the distribution of M-Spike values measured by SPEP are shown (Fig. 1). The residual distribution of the RF model indicated that nearly all M-spike values determined using the 43 variables distributed equally on either side of zero. The addition of race as variables did not significantly change the residual plots. The weighted value of each of the 43 independent variables was determined by individually removing a variable from the ML algorithm and measuring its effect on the mean squared error (MSE). As shown, removal of the first lagged M-spike, serum total protein, second-lagged M-spike, serum IgG, serum IgM, and serum IgA, had the greatest effects on the ML algorithm. The effect of race was then added to the ML algorithm. However, patient race had a low weighted value compared to the other variables included in the ML algorithm (Fig-2). M-spike values determined using the ML algorithm correlated highly with M-spike values determined using the laboratory measured SPEP values as indicated by the proximity of the Pearson and Spearman correlation coefficients to +1. Using the 43 independent variables, the Pearson coefficient was 0.96 and the Spearman coefficient was 0.91, when M-spike values determined using the ML algorithm were compared to laboratory determined M-spike values. When race was added as a variable in the ML algorithm, the Pearson coefficient was 0.95 and the Spearman coefficient was 0.94 which is not significantly different.

Conclusion: Here, we present the results of our machine learning analysis, which further supports our hypothesis and sheds light on the potential racial difference in the overall systemic impact of MM. Understanding the factors underlying racial

disparities in MM will pave the way for more personalized and equitable treatment strategies, ultimately improving patient outcomes and overall disease management.

Disclosures Malek: Karyopharm: Speakers Bureau; Cumberland Inc.: Research Funding; Medpacto Inc.: Research Funding; BMS: Consultancy; Sanofi: Consultancy; Amgen: Speakers Bureau.

Figure-1- Patient-derived M-spike values. Shown is a histogram indicating the number of observations plotted vs. the magnitude distribution for M-spike values among all the patients included in the study.

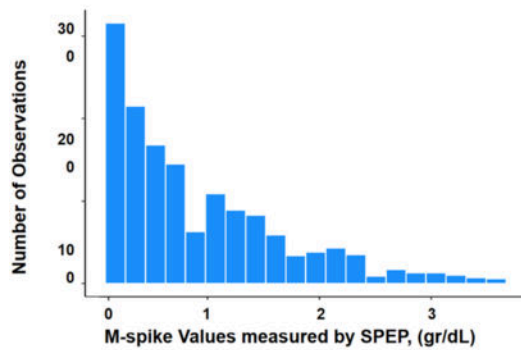
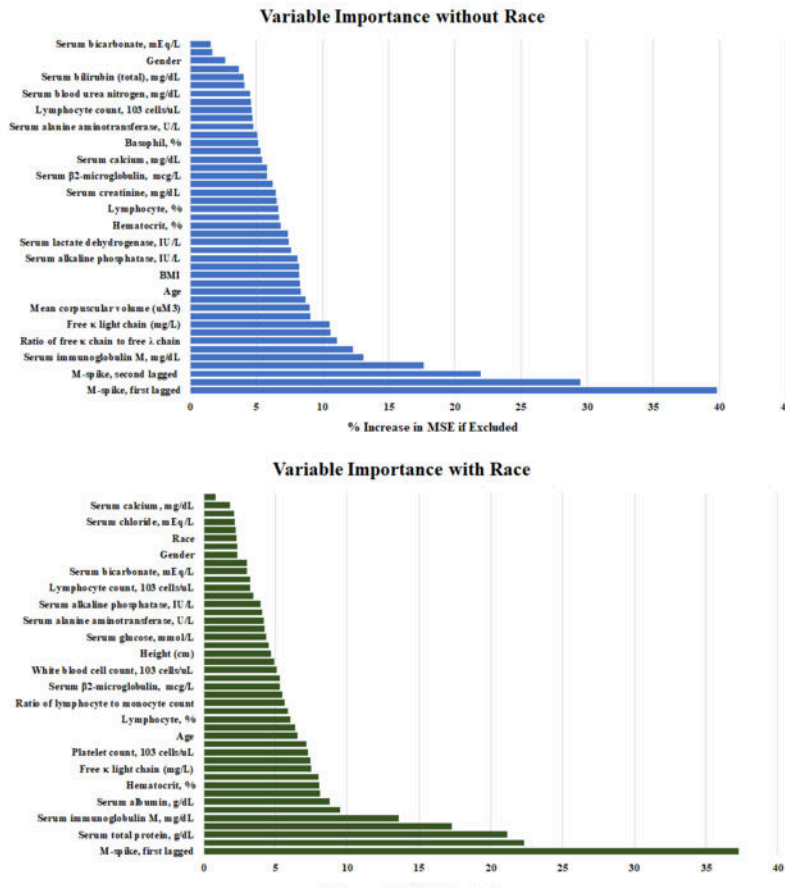


Figure-2- The bar plots illustrating the importance of variables in the two Random Forest (RF) models: one without the inclusion of the Race variable (blue) and the other with the inclusion of the Race variable (green). The importance of a variable determined by how much the model's prediction error increases when the values of that variable are permuted randomly. This is represented as the percentage increase in mean squared error (MSE). A higher percentage increase in MSE indicates that the variable is more important for the model's prediction accuracy.



Downloaded from http://ashpublications.net/blood/article-pdf/142/Supplement_1/7343/20202849/blood-3939-main.pdf by guest on 02 June 2024

Figure 1

<https://doi.org/10.1182/blood-2023-172742>

Downloaded from http://ashpublications.net/blood/article-pdf/142/Supplement_1/7343/2202849/blood-3939-main.pdf by guest on 02 June 2024