

Applying machine learning and text analysis to identify factors that may predict hypertensive heart disease patient outcomes in home healthcare

David Patrishkoff, Nova Southeastern University, USA, dpatrish@nova.edu

Stephen E. Bronsburg, Nova Southeastern University, USA, bronsbur@nova.edu

Mariah Ali, Nova Southeastern University, USA, ali.mariah19@gmail.com

Abstract

This research focuses on predicting the patient discharge disposition with initial patient assessment and therapy data as well as determining which therapy intervention text had positive impacts on hypertension heart disease patients in home healthcare environments. Older adults prefer to stay in their home, which is known as aging in place. Home healthcare is the last line of defense before advancing to other expensive healthcare options. This research used aggregate transactional data from 2,181 home healthcare patients in the United States (U.S.) from 2016-2022. We used the Centers for Disease Control and Prevention (CDC) Patient Driven Groupings Model and focused on the cardiac circulatory patient's subcategory of hypertensive heart disease. Data was analyzed from Activity of Daily Life (ADL) assessment scores, the number of disease diagnosis codes per patient, the number of additional cardiac comorbidities, gender, age, standardized hospitalization risks, number of medications per patient, number of interventions per patient, and the length of stay in home healthcare. Machine learning and advanced text analysis were applied to determine which factors and therapy intervention text had the biggest impact on hypertensive heart disease patient outcomes. This research also identified those interventions with the best Signal to Noise (SN) ratios that are currently being piloted in home healthcare settings.

Keywords: Health informatics, hypertensive heart disease, activities of daily living, therapy interventions, machine learning, text analysis.

Introduction

This study reviewed 2,181 hypertensive heart disease patients in home healthcare settings between 2016 and 2022 that were treated with 1,644 different therapy interventions. Our research data came from RiverSoft Services LLC (RiverSoft, 2022) which is a home healthcare software company who supports healthcare agencies whose patients are located in the central, east, and southeast parts of the United States (U.S.). Hypertensive heart disease occurs over a long period of time. This condition takes years to develop. Patients with high blood pressure (>120/80 mmHg) run greater risks of adverse heart conditions. Heart failure and conduction arrhythmias are probable for those with high blood pressure that is not under control. Heart disease increases for those with high blood pressure that is not under control which may ultimately lead to heart failure especially for those over the age of 65. High blood pressure forces the heart to work harder to pump blood.

This strain on the heart may make the muscle thicken and restrict blood flow. As the heart muscle weakens it can lead to heart failure. High blood pressure will cause the walls of the arteries to thicken and reduce blood flow. Add to this high cholesterol and this puts more stress on the arteries and blood flow. Heart attack and stroke risks increase when arteries and blood flow is restricted and or collects in a stationary manor (clevelandclinic.org, 2022). One of this study's goals was to identify the specific therapy intervention words recorded in medical records that were associated with therapy interventions having positive impacts on hypertension heart disease home healthcare patients. Therapy intervention words were analyzed from 2,181 patients from a national home healthcare database. These intervention words are documented in the home healthcare medical records.

Our patient data for this study included the Activities of Daily Living (ADL) scores which are used as a standard in health care for patient assessments and was an important component of this research. An ADL is defined as "vital functional abilities that relate to self-care" (Margolis et al., 2003). The ADL-Katz ADL index comprises of grooming/bathing or showering, walking, eating, use of stairs, toileting, going outdoors, dressing/undressing, standing up/sitting down (Cabañero-Martínez et al., 2009; Valderrama-Gama et al., 2002). The identification of therapy interventions that are associated with positive ADL outcomes may provide additional valuable insights to improve patient outcomes. This information is intended to augment rather than replace the current factors used for patient care decision making (Ngiam & Khor, 2019). It is also anticipated that other factor correlations discovered in this study can be applied to the home healthcare provider's decision-making process to achieve positive patient outcomes.

According to the World Health Organization (WHO, 2004), coronary artery disease is the leading cause of death worldwide (WHO, 2021), including the U.S. (CDC, 2020). Approximately 700,000 people, or 1 in 4 deaths in the U.S. are due to heart disease (CDC, 2020). The calculated cost of heart disease which includes medication, treatment, lost work productivity in the U.S. is estimated to be \$363 billion yearly (Virani et al., 2021). Experts predict the global population of people 80 and older will more than triple between 2015 and 2050, growing from 126.5 million to 446.6 million. The oldest population in some Asian and Latin American countries will quadruple in the same timeframe. America's 65-and-over population is projected to nearly double over the next three decades, from 48 million to 88 million by 2050. Starting in 2030, when all boomers will be older than 65, older Americans will make up 21% of the population, up from 15% in 2022. (Cire, 2016). The increased numbers of elderly adults pose a challenge to institutions and government resources globally (Partridge et al., 2018). Nine out of 10 Americans 50 and older want to "age-in-place" (Capital Caring Health, 2021). According to Young et al., (2017), choosing the option of home care is linked with a higher quality of life and reduced hospital admissions when compared to institutionalization. The increased use of technology such as mobile medical devices and telemedicine has given leverage to those choosing to age-in-place. Those who are more acceptant of new technologies such as mobile medical devices are more successful with aging-in-place goals (Chimento-Diaz et al., 2022).

With a rapidly aging population and 10,000 people turning 65 each day, the challenge of providing quality and affordable health care to older and disabled Americans has never been more apparent, and the need to meet that challenge has never been more vital. In 2016, there were more than two million personal healthcare workers in home healthcare settings, according to the U.S. Department

of Labor, with the number projected to increase 40% in the next decade (Cheney, 2018). Taking care of elder adults with chronic health conditions and the emotional decisions that must be made when caring for a loved one can be overwhelming. The association between caregivers' emotional stress and care receivers' depression levels exemplifies how navigating later life transactions has an intergenerational and systemic impact on mental wellness (Family Caregiver Alliance, n.d.).

The Centers for Disease Control and Prevention (CDC) (2022) defined aging-in-place as "The ability to live in one's own home and community safely, independently, and comfortably, regardless of age, income, or ability level" (para. 7). Making the decision to age-in-place requires planning and resources. A practical approach such as a community-based care process model includes integration and implementation of daily activities for older adults helping to ensure success (Galof & Balantie, 2021). Troutman-Jordan and Staples (2014) stated that interventions focused on functional performance mechanisms as provided by home healthcare contributes to the quality of life of older adults and promotes successful aging-in-place. Having a means to monitor and review data is a critical factor in patient management. There are several services available that help to collect and store home healthcare data. Medicare (2022) identified home healthcare as "a wide range of health care services that can be given in your home for an illness or injury. Home healthcare is usually less expensive, more convenient, and just as effective as care you get in a hospital or Skilled Nursing Facility (SNF)" (para. 1). Most users of home healthcare are those over age 65. However, according to the National Center for Health Statistics (2019), approximately 18% of the users of home healthcare are under age 65. The average age of patients in our home healthcare dataset was 78.6 years, with an age range of 3 to 109 years. About 60.2% were females and 39.8% were males. These may be individuals that have a health condition requiring home healthcare such as those with disabilities who may also qualify for Medicare or other home healthcare assistance.

Home healthcare plays a leading role in transitioning patients back to their home surroundings with or without support after a health incident such as a discharge from a health care facility or after being treated for a heart attack. Readmission back to a healthcare facility shortly after discharge is a telling sign of health and or care issues. A recent publication by Sheikh et al. (2021) stated that only a small portion of patients released from a hospital after being treated for a severe heart health issue received home healthcare, and those that did had lower 30-day readmission rates. Hospital discharge is a crucial time demanding support for transitional success to the home (Posadas-Collado et al., 2022). There are studies focused on nurse-led transitional care interventions focused on coronary artery patients, however these interventions have yet to be analyzed (Li et al., 2021). In 2019, over 10,000 Medicare-Certified Free-Standing Home Health Agencies (HHAs) were operating in the U.S. that delivered 6.37 million episodes of Medicare Part A and B services to 3.57 million home healthcare patients (AHHQI Home Health Chartbook, 2020). The goals of home healthcare services are to offer a lower cost for healthcare services by allowing the patient to stay in their own home while being treated and rehabilitated. Home healthcare services include treating specific medical conditions, improving wound care, assisting in pain management, assessment, teaching patients about medication management, attempting to improve the quality of life for patients, reducing the amount of hospital Emergency Room (ER) visits, and to reducing their readmission rates to hospitals. The following Research Questions (RQ) are addressed by this study:

-
- RQ1: Can we accurately predict the patient discharge disposition (back to the community with or without assistance or to non-institutional hospice) with only preliminary patient assessment data which includes: the activities of daily living assessment scores, number of disease diagnosis codes per patient, number of cardiac comorbidities, gender, age, and hospitalization risks?
- RQ2: Can we accurately predict the patient discharge disposition with the preliminary patient assessment data which includes: the activities of daily living assessment scores, number of disease diagnosis codes per patient, number of cardiac comorbidities, gender, age, and hospitalization risks, with additional information for the number of medications, number of interventions per patient, and length of stay in the home healthcare program?
- RQ3: Which therapy intervention words from the therapist orders are associated with better patient outcomes?

Methods

This research used aggregate transactional data from 2,181 hypertensive home healthcare patients residing in the United States from 2016-2022. We used the CDC Patient Driven Groupings Model (PDGM) which is used for reimbursement (Centers for Medicare & Medicaid Services Patient-Driven Groupings Model, 2020). PDGM (2020) defined clinical characteristics in the diagnosis reported on a claim for the home healthcare patient. This determines resource allocations. One of the clinical characteristic groups used by the PDGM are clinical groupings. The focus for this study was on home healthcare patients that were diagnosed with cardiac and circulatory health issues with the subcategory of hypertensive heart disease as defined by the ICD-10 codes. The CDC (2022) defined the International Classification of Disease, tenth edition (ICD-10) as the tenth version of codes for medical diagnosis and classification of disease, injuries, and health issues including claims and billing. A fair amount of data wrangling and feature engineering (Carver, 2017; Osborne, 2013) was conducted to make the data ready for analysis and to create new calculated features such the sum of nine specific ADL starting scores, length of stay, the number of ICD-10 codes per patient (nine max), the sum of actual hospitalization risks per patient (nine max), the sum of additional cardiac comorbidities (eight max), the number of medications per patient, and the number of therapy interventions per patient.

The data used for this analysis was provided in March 2022 from RiverSoft Services LLC (RiverSoft, 2022). Due to the inherent high quality of their data used for billing purposes, any data issues we experienced were more related to the conversion of that system data into clean exportable data sets for analysis purposes and not necessarily related to data errors in the native system. Nevertheless, this data set and any large data set intended for research and analysis still needs to be rigorously scrutinized, cleaned, scoped, stratified, and strategized for how to deal with suspect data, missing data, and “other” data fields. Substantial feature engineering (discretization, categorical encoding, feature splitting, variable transformation, scaling, and newly created features) was also applied to create new calculated features, data bins, data transformations, and joining with other data sets in different relational databases. No identifiable patient data was transferred to us. Anonymous patient numbers were used in the data set so that we could track which therapy interventions were applied to different patients and when they were applied. Every new data export to us from RiverSoft requires the same amount of data wrangling work with new

challenges that need to be overcome. Any questionable data fields, such as 1/1/1900 start of care and or death dates were deleted with no data imputation applied. At the start of 2019, smoking and obesity status reporting was mostly discontinued in our data sets, so we excluded those factors from the analysis due to the large time gap in the data for those factors. Patients who died were excluded from this study since death dates were not always available and reliable end of care data was questionable. Machine Learning (ML) modeling to predict discharge disposition excluded patients who had a “moved” or “unknown” discharge disposition. Racial identifiers for patients were not applied to this study since 93% of the data was from patients identified as Caucasians. Patient location or zip code analysis was not included in this research due to the large gaps in location data for parts of a specific state or for the whole of the U.S. for which RiverSoft does not have client data. Several continuous factor values were grouped (binned) at times for a simplified representation of the analysis results.

Data Analysis and Results

Table 1 offers a general and simple rank of each studied factor to first show which factors had the highest level of variation when using the ADL Improvement factor as our main metric of performance (Mlinac, 2016).

Table 1. Rankings for Patient Specific Factors that Affect ADL Improvement

Rank	Difference in ADL Improvement mean between the highest and lowest group or bin average	Factor	# of Groups or bins	p-value	How does it correlate with ADL Improvements?	Starting Input Assessment, Therapy, or Care Outcome
1	17.0	ADL Starting Scores for 9 ADLs	40	<.001	Steady Increase, then some decline when compared with ADL improvement scores	Starting Assessment
2	12.6	# of Interventions per patient - Binned	12	<.001	Decreases, when compared to ADL improvements	Therapy Interventions
3	11.3	611 Individual Therapy Interventions*	611	<.001	See SN ratio and therapy intervention text analysis for details	Therapy Interventions
4	11.0	Age - Binned	16	<.001	Increases, then decreases, when compared to ADL improvements	Starting Assessment
5	8.4	Discharge Disposition - To either Hospice, the community with assistance, or to the Community	3	<.001	Hospice patients have the lowest ADL improvements for each ADL starting score	Outcome
6	6.0	# of ICD10 codes per Patient up to 9 codes	9	<.001	Drops dramatically after 1 and then steadily increases	Starting Assessment
7	5.6	# of Medications per Patient - Binned	12	> 0.05	Increases, then is volatile, when compared to ADL improvements	Starting Assessment
8	5.0	Starting assessment for 9 Hospitalization Risks (yes / no)	9	<.001	Positive correlation with ADL improvement scores	Starting Assessment
9	3.0	Additional cardiac comorbidities: Dx2 to Dx9 first letter ICD10 codes	22	> 0.05	The highest and lowest groups have overlapping confidence intervals	Starting Assessment
10	1.4	Length of Stay - Binned	5	<.001	Increases, then drops	Therapy Interventions
11	0.5	Gender	2	<.001	Female ADL improvement scores are higher than males	Starting Assessment

*These individual interventions have been applied at least 10 times

Table 2. M1033 Hospitalization Risks Used for Patient Assessments

1. History of falls (two or more falls – or any fall with an injury – in the past 12 months)	4. Multiple emergency department visits (two or more) in the past six months	7. Currently taking five or more medications
2. Unintentional weight loss of a total of 10 pounds or more in the past 12 months	5. Decline in mental, emotional, or behavioral status in the past 3 months	8. Currently reports exhaustion
3. Multiple hospitalizations (two or more) in the past six months	6. Reported or observed history of difficulty complying with any medical instructions (for example, medications, diet, exercise) in the past three months	9. Other risk(s) not listed in 1 – 8

Table 2 shows the nine standard home healthcare hospitalization risk assessments that were provided for each patient. These are yes or no questions that were assessed at the start of home

healthcare. This research looked at each risk individually as well as the simple sum of the risks that were answered as a “yes” for each patient.

Table 3 shows the nine functional Impairment assessments for specific ADLs. The lowest score in each ADL category represents no functional impairment or pain at all. Some ADLs have a scale of zero to three and others may have an assessment score that ranges as high as zero to six. The highest score for an individual ADL always represents a situation where the patient depends entirely on someone else to conduct that activity and or experiences constant pain. These ADL scores are established at the start and end of a home healthcare regiment of services. The goal of any series of therapy interventions is to improve the ADL scores by the end of the care period while dealing with specific primary diseases that are usually complicated by the presence of other comorbidities. These starting, ending, and resulting improvement of ADL scores were highly integrated in our analysis.

Table 3. Nine of the ADL Patient Assessments Used at the Start and End of Care

M1800 Grooming (0-3)	M1830 Bathing (0-6)	M1860 Ambulation and locomotion (0-6)
M1810 Current ability to dress upper body (0-3)	M1840 Toilet transferring (0-4)	M1870 Feeding / eating (0-5)
M1820 Current ability to dress lower body (0-3)	M1950 Transferring (0-5)	M1242 Frequency of Pain (0-4)

Figure 1 shows that for most ADL starting assessment scores, patients that are discharged to the community, after the care period, experience the highest level of ADL score improvement. Patients returning to the community with assistance have slightly less improvements, and those patients who were discharged to non-institutional hospice saw the lowest levels of ADL improvement during home healthcare treatment programs. These differences in ADL improvement levels between the three discharge disposition groups are statistically significant ($p < 0.05$).

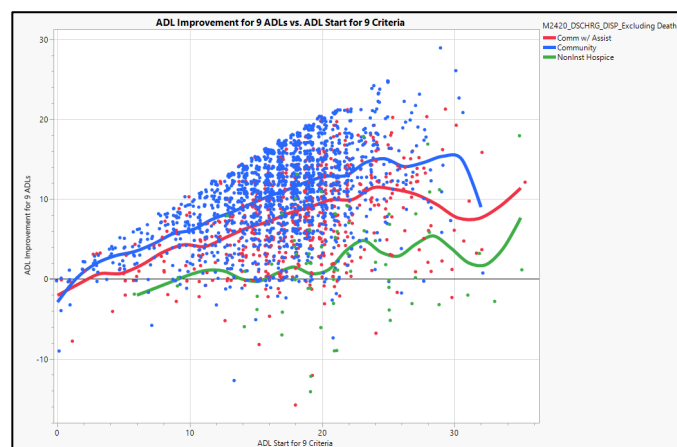


Figure 1. The Average Improvements of ADL Scores Versus the ADL Starting Scores

Figure 2 shows that no matter what sum of the M1033 hospitalization risks are at the start of patient care, the ADL improvements are consistent for each of the 3 different discharge disposition groups: Discharged to the community, with or without assistance, and discharged to a non-institutional hospice. Patients returning to the community experience the greatest ADL improvements while patients who are discharged to non-institutional hospice see the lowest levels of improvement

during home healthcare treatment programs. These differences in ADL improvement levels between the three discharge disposition groups are statistically significant ($p < 0.05$).

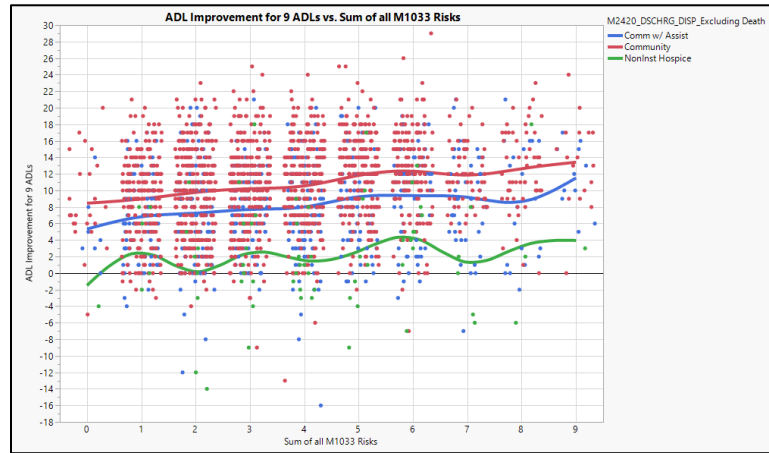


Figure 2. The Average Improvements of ADL Scores Versus the Sum of All Hospitalization Risks

Increases in the number of therapy interventions per patient is no guarantee that ADL scores will improve at the end of a care period. Figure 3 shows that each of the three different discharge disposition groups shows a point of diminishing returns for increased numbers of interventions per patient. No matter what the number of interventions per patient is, patients discharged to the community always experience the most improvements with patients that are discharged to non-institutional hospice experience the lowest levels of ADL improvements, no matter how many therapies they are exposed to. These differences in ADL improvement levels between the three discharge disposition groups are statistically significant ($p < 0.05$).

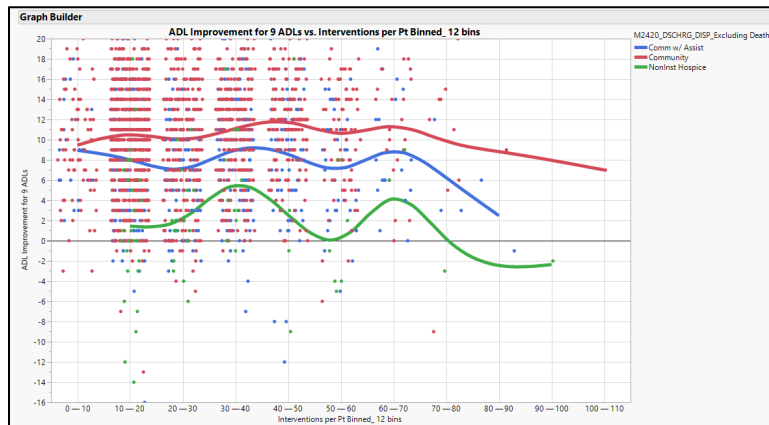


Figure 3. Interventions Per Patient Compared to ADL Improvement Levels

Principal Component Analysis (PCA), Factor Analysis, column clustering, and 2-way hierarchical clustering were conducted to explore dimensionality reduction opportunities and the identification of factor clusters for both of the models mentioned in RQ1 and RQ2. The PCA analysis charts shown in Figure 4 refers to RQ1. Individual clustering analysis results varied but some common themes agreed with each other, such as the generally agreed on the importance for the nine ADL

starting scores, followed by multiple recent ER visits and hospitalizations, followed by other hospitalization risk groupings, and other factors. Figure 4 shows some of the PCA analysis charts, including the partial contributions of variables chart and the Eigenvalues chart which shows us, among other things, that the top 13 rated factors represent roughly 80% of variation found in all 22 of the initial patients starting assessment factors.

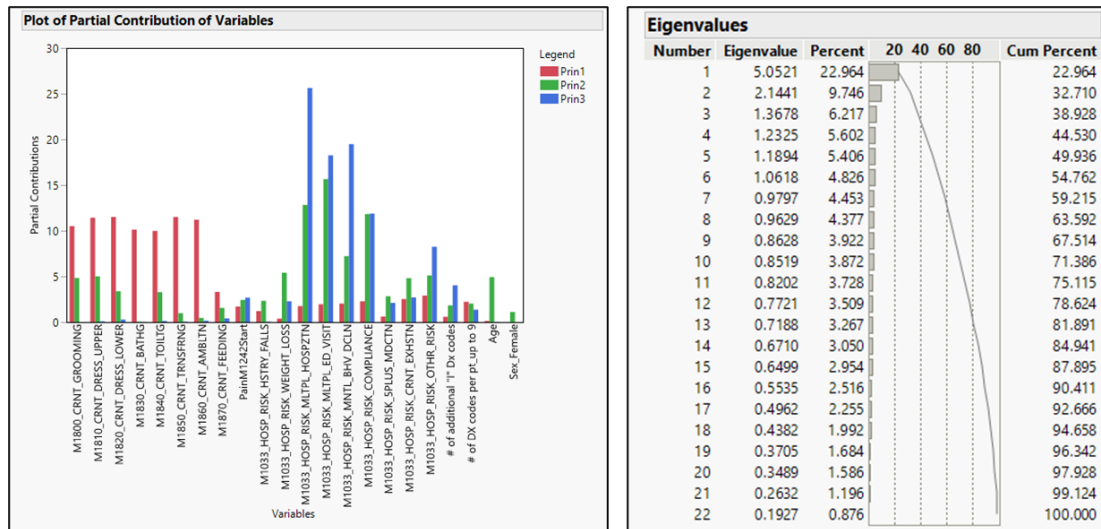


Figure 4. Principal Components Analysis (PCA) for 22 Initial Patient Assessments

Machine Learning Analysis

ML model screening and final model building was conducted with JMP Pro software (Barker et al., 2021) to identify the best model for two areas of interest:

1. Predicting patient discharge disposition (back to the community with or without assistance or to a non-institutional hospice) with just the initial patient assessment data: ADL starting scores for each of the nine ADL assessments, the number of interventions per patient, age, number of ICD-10 codes per patient (nine max), the presence of each of the nine possible hospitalization risks, the number of additional cardiac comorbidities per patient, and gender (See Figure 5 for model results).
2. Predicting patient discharge disposition with the initial patient assessment data with additional information collected during the home health care of patients: the number of interventions per patient, number of medications per patient, and length of stay (See Figure 6 for model results).

Figure 5 shows the results of a machine learning neural network model that used 33% of the data for five-fold cross-validation and activated three hidden layers in the neural network. The neural network algorithm was chosen after initial model screening showed that the neural network model had a 5% better Area Under the Curve (AUC) metric shown on the ROC chart when compared to the 2nd best model and a 7.5% advantage compared to the worst competing model. The other ML models that were considered were as follows, listed in their rank behind the neural network model: Nominal logistic, generalized regression lasso, bootstrap forest, decision tree, XGBoost, and support vector machines. The neural network model was still the best choice even after some

hyperparameter tuning was explored for each of the mentioned models. The covariates in the neural network model were transformed in the final neural network model analysis to address multicollinearity and to improve the convergence to the optimization algorithm.

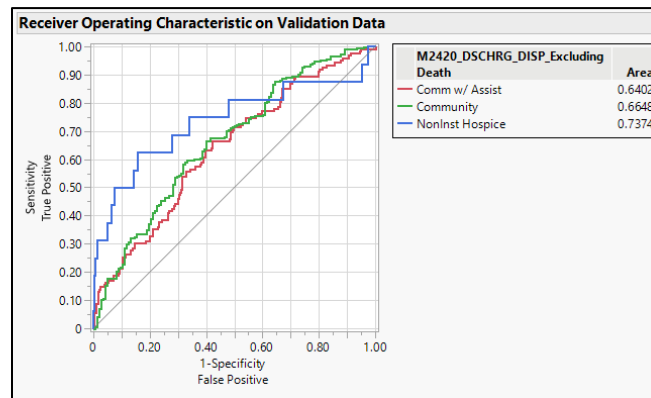


Figure 5. ROC Curves Using Just Initial Patient Assessment Data for Predictions

Figure 6 shows the results of a machine learning neural network model that used 33% of the data for five-fold cross-validation and activated three hidden layers in the neural network. The neural network algorithm was chosen after initial model screening showed that the neural network model had a 1.5% better Area Under the Curve (AUC) metric shown on the ROC chart when compared to the 2nd best model and a 4.8% advantage compared to the worst competing model. The other ML models that were considered were as follows, listed in their rank behind the neural network model: Bootstrap Forest, generalized regression lasso, nominal logistic, decision tree, support vector machines, and XGBoost. The neural network model was still the best choice even after some hyperparameter tuning was explored for each of the mentioned models. The covariates in the neural network model were transformed in the final neural network model analysis to address multicollinearity and to improve the convergence to the optimization algorithm.

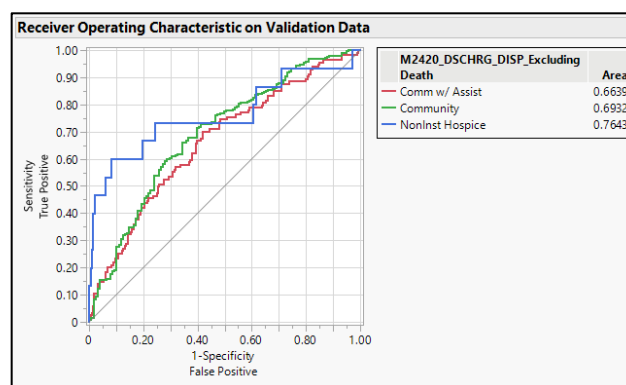


Figure 6. ROC Curves with Patient Starting Assessments and Patient Care Data for Predictions

A ROC chart is a graphical representation of the performance of a classification model at different thresholds for the above-mentioned ML models. The x-axis on a ROC chart represents False Positive Rate (FPR) and the y-axis represents True Positive Rate (TPR). A ROC curve is generated by plotting the TPR against the FPR at different thresholds. In general, a good AUC rate for a ROC

chart is typically considered to be above 0.7, which was achieved for the non-institutional hospice discharge group. Continued follow-up research will add other factors and clusters to this model in attempts to improve the accuracy of this predictive model for all discharge dispositions.

Decision trees (Rokach & Maimon, 2015) are not always the most accurate algorithms in the machine learning arsenal, but they offer highly explainable results that almost anyone can understand. They can be tuned to either show very simple high-level results, optimal accuracy levels with more complexity, or very detailed versions of the tree chart with over-fitted results. Figure 7 shows an optimal accuracy level decision tree created with 20% validation data in BlueSky Statistics software (Muenchen, 2021), based on R packages. Such decision trees can be used as the basis for compelling conversations between different healthcare professionals and to build levels of basic understandings in machine learning tools amongst healthcare professionals. In general, the right side of the tree showed decision tree splits that yielded high ADL improvements while the left side of the decision tree showed decision tree splits that yielded lower ADL improvements.

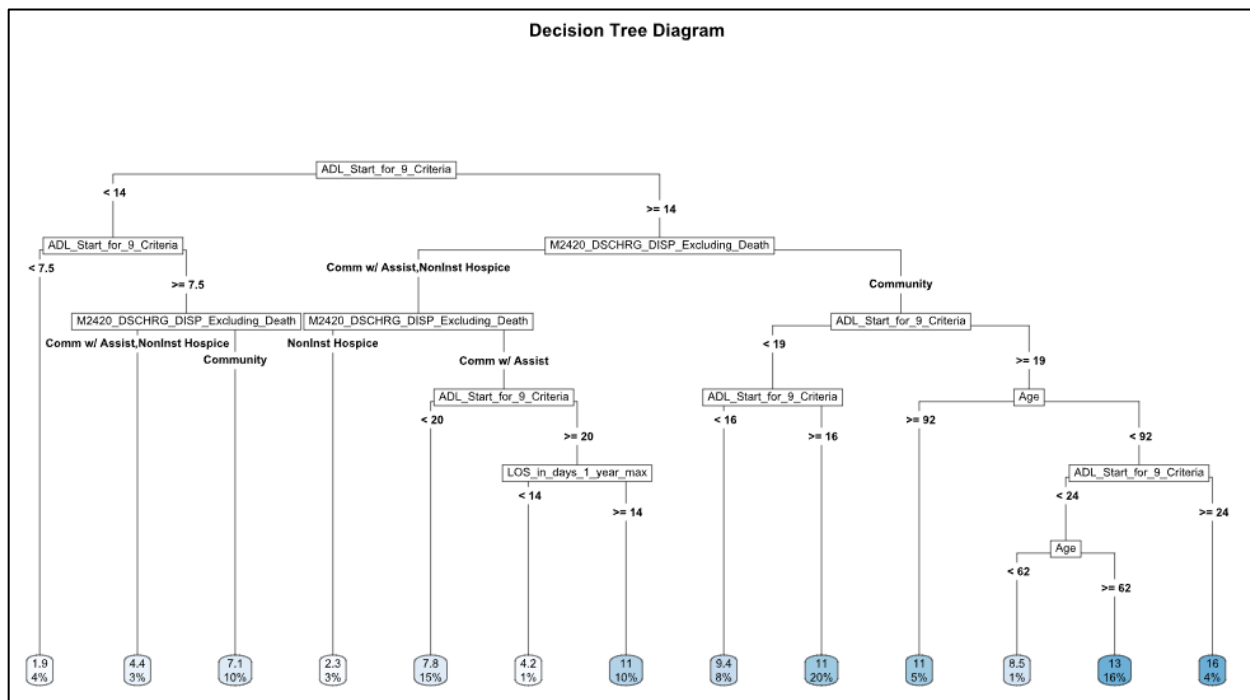


Figure 7. A Machine Learning Decision Tree with the Factors That Affect ADL Improvement Levels

Selecting the Best Therapy Interventions

This research investigated various ways to select the best performing hypertensive interventions out of the total of 1,644 hypertensive interventions used in the data base. Unfortunately, intervention wording is not always consistent within or between home healthcare agencies. If one of the interventions has one different word in the description, it will be counted as a different intervention by JMP and with other text explorer analysis packages. On average, 25 words were used to describe a therapy intervention for hypertensive care. We applied a filtering method to

only analyze those interventions that were used at least 10 times during our analysis scope. This ten-intervention usage minimum guideline was determined and agreed upon with RiverSoft clinical experts as an appropriate filter. They did not want the outlier performance of interventions with a small number of usages to drive us to an unstable intervention recommendation. This filter narrowed our deeper analysis to only 611 interventions. Then, we calculated the mean ADL Improvement for each intervention, which we call the signal. We then calculated the standard deviation for each ADL improvement, which we called noise. We divided the signal values by the noise values to create a Signal to Noise (SN) ratio metric that was useful for intervention scoring. The higher the SN ratio for an intervention, the more it had the potential to improve the ADL scores with the least amount of variation in its performance. Figure 8 shows the performance of the interventions that were binned into various SN ratio ranges with the visualization application of notched box plots. The notches on the box plot identify the confidence intervals for the medians. The white horizontal line in the box is the median. Jittered dots above and below the boxes identify outliers. The N-values represent the number of different interventions represented for each SN ratio bin.

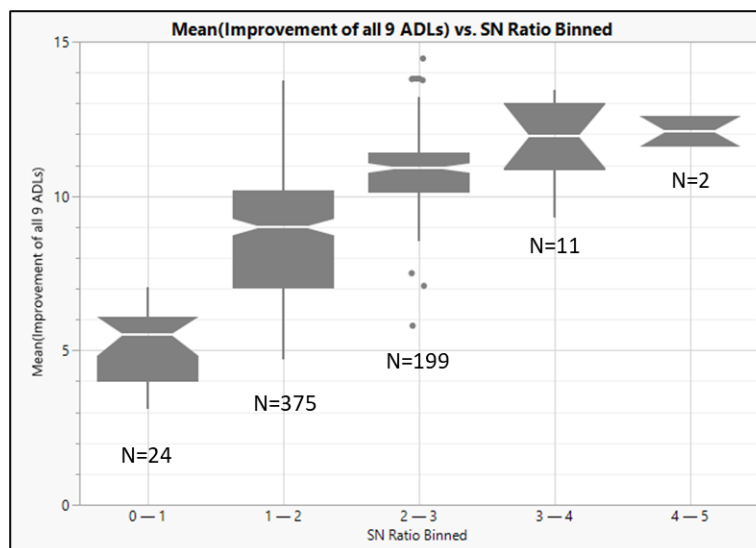
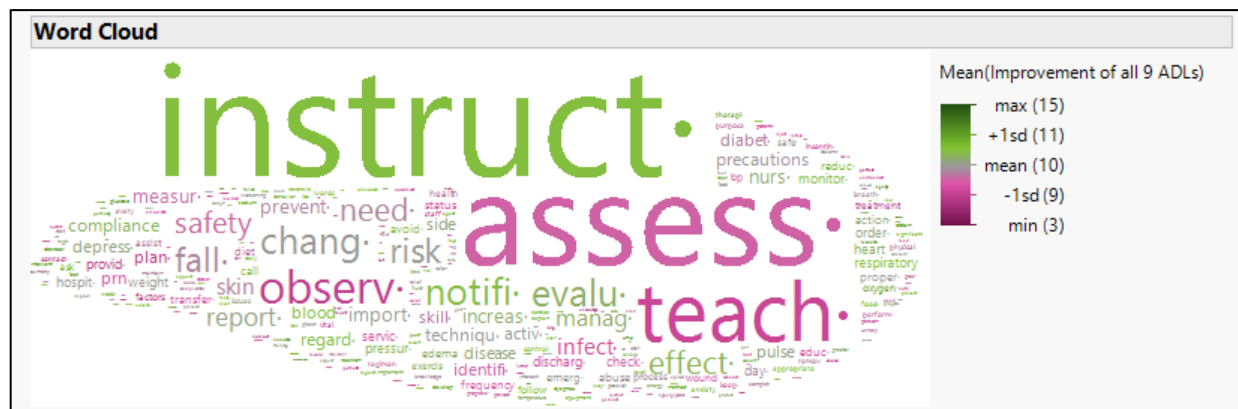


Figure 8. Signal to Noise (SN) Ratios for All Interventions That Were Used at Least 10 Times

Additional intervention text analysis was performed to check if the presence of certain terms in the intervention text represented higher performing interventions. The JMP Pro text explorer analysis capabilities (Ingersoll, 2021) were used to support this analysis. Figure 9 shows a word cloud with the word stemming feature which puts a dot behind the word to represent that a core stem word was used in the frequency analysis. A stem word is a part of a word that has several possibly continuation options for that word. Word Clouds (Hamm, 2011) are popular tools to identify high frequencies of a specific word or term in text with larger font sizes to show more frequently used words. We applied an additional advanced JMP word cloud feature that also color codes the words in the word frequency cloud to represent a higher or lower average ADL improvement score for that word. The word cloud in Figure 9 adds the before-mentioned feature while preserving the classic word cloud feature where the font size correlates with word usage frequencies. Rose colored text in the word cloud in Figure 9 represents text that was used in



Singular Value Decomposition (SVD) (SAS, 2021) is a mathematical way of breaking down large amounts of interventions texts into smaller latent groups of important topics or themes. Figure 10 shows just part of the SVD analysis that was carried out on the intervention text in this project. This SVD analysis also offered the top topic loading values for each word used in the word clouds that this analysis created for each of the 10 latent topic groups that were identified.

This SVD analysis of the intervention text helped us to discover 10 different topic clusters for all intervention texts. Careful analysis of the SVD analysis outputs allowed us to create a meaningful definition for each intervention cluster topic group with the highest explained variance topics listed first on the list below.

- 36 -

7. Assess the effects of diabetes and instruct on foot care.
8. Assess and instruct in fall prevention.
9. Teach about appropriate activity limitations, restrictions, and the importance of rest periods.
10. Instruct on the importance of maintaining a stable weight and a low sodium diet.

Future Research

Word cloud analysis will be expanded to include intervention phrases, text principal components, and common intervention topics to discover the most frequent and high performing intervention phrases, themes, and topics that were used. A basic challenge encountered in this research project was that many different interventions were applied to each patient which makes the assessment of individual intervention contributions difficult. Association analysis will be applied to help identify the most frequent and best performing groups of interventions (Liu & Xu 2016). Further work will be continued to pursue high-performance ML ensemble models to attempt an improvement in the predictive power of the current ML models. Sentiment analysis (Zhao, 2020), latent semantic analysis (Landauer et al., 2007) and term selection (Baker et al., 2021) will be continued to gain additional insights. Ongoing research will include the integration of the following features to improve the scope and accuracy of future ML models: individual patient medications with dosages, individual ICD-10 codes of comorbidities, and individual interventions. The integration of these newly mentioned model features in the analysis will be achieved with the help of indicator variables, subsequent clustering, and latent group analysis. We will also continue to analyze home healthcare data from those patients who died to see if any insights can be gained from that analysis in spite of its gaps in analyzable data.

Discussion and Conclusion

Our initial analysis and machine learning models offer hope that we will be able to improve the predictions for discharge dispositions with starting patient assessments as we continue to integrate other factors in this ongoing research. SN ratio scoring for therapy interventions offers a simple way to score the effectiveness of interventions but that analysis does not tell us why a specific therapy intervention was successful or not. Additional text analysis tools have helped us to answer that “why” questions. One analysis challenge is that therapy intervention wording is not always consistent between and within home healthcare agencies. However, common words exist within and between different interventions that can be analyzed. Word clouds with additional color-coding for intervention performance has offered greater insights into which words in a therapy intervention are associated with better patient outcomes. Singular value decomposition analysis of intervention text has also been very useful in determining the top 10 themes for all interventions which will be useful in further analysis of the clustered intervention text. The home healthcare sector currently relies too much on well-intended individual instincts and experience levels to make patient intervention decisions that greatly influences the quality of life for their patients. These decisions are made without seeking the wisdom that is hidden in larger national home healthcare databases that can offer valuable evidence-based decision-making insights. The application of these research findings is currently being tested in the real world. RiverSoft is offering home healthcare nurses real-time access to high SN ratio interventions that will be represented in our expanded research that applies to all clinical groups and ICD-10 disease codes.

Such activities are in the initial pilot phase to offer nurses real-time recommendations based on our wide range of evidence from the field to augment the expertise and experiences of nurses with silent second opinions. The final decision makers for intervention usages are physicians and the actual home healthcare providers of care to patients but we are hopeful that real time access for these professionals to the best patient interventions can improve the final decision-making process. We are also hopeful that further improvements in the accuracy of our predictive discharge disposition models will be able to provide home health care professionals with early warning signals for non-optimal patient outcomes that should be further assessed and addressed.

References

- AHHQI Home Health Chartbook. (2021). *Alliance for home health quality and innovation*. <https://www.ahhqi.org/research/home-health-chartbook>
- Barker, C., Falson, P., & Pasour, E. (2021). A prediction is worth a thousand words: An introduction to term selection in JMP® Pro (2021-EU-30MP-747). *JMP Statistical Recovery*. <https://community.jmp.com/t5/Discovery-Summit-Europe-2021/A-Prediction-Is-Worth-a-Thousand-Words-An-Introduction-to-Term/ta-p/349214>
- Carver, R. (2017). *Preparing data for analysis with JMP*. SAS Institute Inc.
- Cabañero-Martínez, M. J., Cabrero-García, J., Richart-Martínez, M., & Muñoz-Mendoza, C. L. (2009). The Spanish versions of the Barthel index (BI) and the Katz index (KI) of Activities of Daily Living (ADL): A structured review. *Archives of Gerontology and Geriatrics*, 49(1), e77–e84. <https://doi.org/10.1016/j.archger.2008.09.006> .
- Capital Caring Health, (2021). Nearly 90% of Americans age 50 and older want to “Age in Place”. *Global Newswire*. <https://www.globenewswire.com/news-release/2021/05/10/2226492/0/en/NEARLY-90-OF-AMERICANS-AGE-50-AND-OLDER-WANT-TO-AGE-IN-PLACE.html>
- Centers for Disease Control and Prevention (CDC) (2020). Underlining cause of death, 1999-2020. *CDC WONDER Online Database*. <https://wonder.cdc.gov/ucd-icd10.html>
- Center for Disease Control and Prevention (CDC) (2022). Healthy places terminology: Aging in place. <https://www.cdc.gov/healthyplaces/terminology.htm>
- Centers for Medicare and Medicaid Services: Medicare Learning Network, (2020). Overview of the patient-driven groupings model. *MLN Matters NumberSE19027*. <https://www.cms.gov/files/document/se19027.pdf>
- Chimento-Díaz, S., Sánchez-García, P., Franco-Antonio, C., Santano-Mogena, E., Espino-Tato, I., & Cordovilla-Guardia, S. (2022). Factors associated with the acceptance of new technologies for ageing in place by people over 64 Years of Age. *The International Journal of Environmental Research and Public Health*, 19(5), 2947. <https://doi.org/10.3390/ijerph19052947>

-
- Cheney, C (2018). 5 guiding principles to improve home health safety. *Healthleaders*.
<https://www.healthleadersmedia.com/clinical-care/5-guiding-principles-improve-home-health-safety>
- Cire, B. (2016). World's older population grows dramatically. *National Institute of Health Turning Discovery Into Health*. <https://www.nih.gov/news-events/news-releases/worlds-older-population-grows-dramatically>
- Cleveland Clinic (2022). Hypertensive heart disease. *Cleveland Clinic*.
<https://my.clevelandclinic.org/health/diseases/21840-hypertensive-heart-disease>
- Family Caregiver Alliance, (2022). *Taking care of YOU: Self-care for family caregivers*.
<https://www.caregiver.org/resource/taking-care-you-self-care-family-caregivers/>
- Galof, K., & Balantić, Z., (2021). Making the decision to stay at home: Developing a community-based care process model for aging in place. *The International Journal of Environmental Research and Public Health* 18(11), 5987.
<https://doi.org/10.3390/ijerph18115987>
- Hamm, S. (2011). Using word clouds for reflection and discussion. *Journal of Teaching Theology and Religion*, 14(2), 156. <https://doi.org/10.1111/j.1467-9647.2011.00692.x>
- Ingersoll, G., Morton, T., S., & Farris, A., L. (2012). *Taming text: how to find, organize, and manipulate it*. Manning Publishing Co.
- Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). *Handbook of latent semantic analysis*. Psychology Press, Taylor & Francis Group.
- Li, M., Li, Y., Meng, Q., Li, Y., Tian, X., Liu, R., & Fang, J. (2021). Effects of nurse-led transitional care interventions for patients with heart failure on healthcare utilization: A meta-analysis of randomized controlled trials. *PLoS One*, 16(12), e0261300.
<https://pubmed.ncbi.nlm.nih.gov/34914810/>
- Margolis, S. A., Carter, T., Dunn, E. V., & Reed, R. L. (2003). Validation of additional domains in activities of daily living, culturally appropriate for Muslims. *Gerontology*, 49(1), 61–65. <https://doi.org/10.1159/000066509>
- Merriam Webster (2022). *Instruct*. <https://www.merriam-webster.com/dictionary/instruct>
- Medicare (2022). What Medicare covers. What's home health care?
<https://www.medicare.gov/what-medicare-covers/whats-home-health-care>
- National Center for Health Statistics (2019). Long-term care providers and services used in the United States, 2015-2016. *Vital and Health Statistics*. U.S. Department of Health and Human Services. https://www.cdc.gov/nchs/data/series/sr_03/sr03_43-508.pdf
- Mlinac, M. E., & Feng, M. C., (2016). Assessment of activities of daily living, self-care, and independence. *Archives of Clinical Neuropsychology*, 31(6), 506-516.
<https://doi.org/10.1093/arclin/acw049>
- Muenchen, R. (2021). BlueSky statistics 10.0 user guide: 1st draft. *Lulu*. www.lulu.com

-
- Osborne, J., W. (2013). Best practices in data cleaning. *Newborn and Infant Nursing Reviews*, 10(1).
- Ngiam, K. Y., & Khor, I. W. (2019). Digital oncology: Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology*, 20(5), e262-e273.
[https://doi.org/10.1016/S1470-2045\(19\)30149-4](https://doi.org/10.1016/S1470-2045(19)30149-4)
- Partridge, L., Deelen, J., & Slagboom, P.E. (2018). Facing up to the global challenges of ageing. *Nature*, 561, 45–56. <https://doi.org/10.1038/s41586-018-0457-8>
- Posadas-Collado, G., Membrive-Jiménez, M.J., Romero-Béjar, J.L., Gómez-Urquiza, J.L., Albendín-García, L., Suleiman-Martos, N., & Cañadas-De LaFuente, G.A. (2022). Continuity of nursing care in patients with coronary artery disease: A systematic review. *International Journal of Environmental Research and Public Health*, 19(5), 3000.
<https://doi.org/10.3390/ijerph19053000>
- RiverSoft Services LLC. (2022). *Unpublished raw Outcome Assessment Information Set (OASIS) data*, Melbourne, FL.
- Rokach, L. & Maimon, O. (2015). *Data mining with decision trees: Theory and application*. World Scientific Publishing Co., Inc.
- Sheikh, M.A., Ngendahimana, D., Deo, S.V., Raza, S., Altarabsheh, S., Reed, G.W., Kalra, A., Kapadia, S.R., & Eagle, K. (2020). Home health care after discharge is associated with less early readmissions for patients with acute myocardial infarction. *Cardiovascular Quality and Outcomes* 13(1), A107. https://doi.org/10.1161/hcq.13.suppl_1.107
- Statistical Analysis System Institute Inc. (2022). *JMP® 16 documentation library*. SAS Institute Inc.
- Troutman-Jordan, M., & Staples, J., (2014). Successful aging from the viewpoint of older adults. *Research and Theory for Nursing Practice: An International Journal*, 28(1).
<https://doi.org/10.1891/1541-6577.28.1.87>
- Valderrama-Gama, E., Damián, J., Ruigómez, A., & Martin-Moreno, J. M. (2002). Chronic disease, functional status, and self-ascribed causes of disabilities among noninstitutionalized older people in Spain. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, 57(11), M716–M721.
<https://doi.org/10.1093/gerona/57.11.M716>
- Virani, S. S., Alonso, A., Aparicio, H. J., Benjamin, E. J., Bittencourt, M. S., Callaway, C. W., Carson, A. P., Chamberlain, A. M., Dellings, F. N., Elkind, M. S. V., Evenson, K. R., Ferguson, J. F., Gupta, D. K., Khan, S. S., Kissela, B. K., Knutson, K. L., Lee, C. D., Lewis, T. T., Liu, J., Loop, M. S., ... Tsao, C. W. (2021). Heart disease and stroke statistics-2021 update. *American Heart Association Council on Epidemiology and Preventative Statistics Committee and Stroke Statistics Subcommittee*.
<https://doi.org/10.1161/CIR.0000000000000950>
- World Health Organization (2021). *Cardiovascular disease (CVD)*. https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1

World Health Organization (2004). A glossary of terms for community health care and services for older persons. *WHO Center for Health Development*.

<https://apps.who.int/iris/handle/10665/68896>

Young, C., Hall, A. M., Gonçalves-Bradley, D. C., Quinn, T. J., Hooft, L., van Munster, B. C., & Stott, D. J. (2017). Home or foster home care versus institutional long-term care for functionally dependent older people. *Cochrane Database Systems Review*, 4(4), CD009844. <https://doi.org/10.1002/14651858.CD009844.pub2>

Zhao, J., Liu, K., & Xu, L. (2016). Sentiment analysis: Mining opinions, sentiments, and emotions. *Computational Linguistics*, 42(3), 595–598.

https://doi.org/10.1162/COLI_r_00259

Authors Biographies

David Patrishkoff, M.S. is a Lean Six Sigma Master Black Belt with C-level worldwide Executive experiences at multi-billion-dollar revenue companies. He is best known for his ability to rapidly solve mission-critical business problems and for his analytical and machine learning skills as they apply to business problem-solving and for the analysis of large home healthcare data sets. He has led many successful business transformations and data analysis projects for healthcare, service, and manufacturing organizations from over 60 different industries worldwide. He also teaches various classes as an Adjunct Professor at the Kettering University School of Management in Flint Michigan, and at the Dr. Kiran C. Patel Osteopathic School of Medicine in Ft. Lauderdale, Florida, which is part of Nova SE University. Since 2017, David has included machine learning (ML) and AI analysis techniques to his consulting work, teachings, and research. He has spoken at many international conferences with a special focus on the application of no-code ML and AI analysis techniques in recent years. David has trained, certified, and mentored over 3,000 professionals in Lean Six Sigma and trained over 23,000 healthcare professionals in High Reliability Organizations (HRO) techniques to reduce medical errors in healthcare systems.



Stephen E. Bronsburg, Ph.D., M.S., M.H.S.A. is the Interim Chair of the Department of Health Informatics, Director of the Master of Science in Health Informatics Program, and Associate Professor at Nova Southeastern University (NSU) Dr. Kiran C. Patel College of Osteopathic Medicine (KPCOM) Fort Lauderdale, Florida, USA. He is a Peer Reviewer for Commission on Accreditation for Health Informatics and Information Management Education (CAHIIM), and a board member for the South Florida Healthcare Information and Management Systems Society (SFHIMSS). Bronsburg conducts research of health information technology (HIT) with an emphasis on HIT implementation, application and evaluation focusing in the areas of health care delivery interventions addressing chronic disease and addiction. Bronsburg has authored and coauthored numerous peer reviewed journal articles, conference proceedings, and presentations nationally and international. Bronsburg is a peer reviewer for the American Medical Informatics Association (AMIA), sits on the Editorial Review Board for the



Online Journal of Applied Knowledge Management as well as a peer reviewer for the International Institute for Applied Knowledge Management Conference.

Maria Ali, M.S. is a Healthcare informatics graduate from Nova Southeastern University and a Six Sigma Blackbelt certified professional. She has a background in manufacturing operations management with a focus on process improvement. She is currently working with Kaiser Permanente as a Business Analyst and e-commerce Product platform owner. she aims to provide solutions and strategies to improve organizational performance and productivity with her extensive knowledge in healthcare IT and software development life cycles. Her career goals are to use her data analysis skills and machine learning techniques to improve the efficiency of healthcare systems.

