MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190310

Using Machine Learning on Home Health Care Assessments to Predict Fall Risk

Yancy Lo[a], Selah F. Lynch[a], Ryan J. Urbanowicz[a,b], Randal S. Olson[b], Ashley Z. Ritter[c], Christina R. Whitehouse[c,d], Melissa O'Connor[d], Susan K. Keim[c], Margaret McDonald[e], Jason H. Moore[a,b], Kathryn H. Bowles[c,e]

[a] *Clinical Research Informatics Core, Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA, USA*
[b] *Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA, USA*
[c] *NewCourtland Center for Transitions and Health, School of Nursing, University of Pennsylvania, Philadelphia, PA, USA*
[d] *Fitzpatrick College of Nursing, Villanova University, Villanova, PA, USA*
[e] *Center for Home Care Policy & Research, Visiting Nurse Service New York, New York, NY, USA*

Abstract

Falls are the leading cause of injuries among older adults, particularly in the more vulnerable home health care (HHC) population. Existing standardized fall risk assessments often require supplemental data collection and tend to have low specificity. We applied a random forest algorithm on readily available HHC data from the mandated Outcomes and Assessment Information Set (OASIS) with over 100 items from 59,006 HHC patients to identify factors that predict and quantify fall risks. Our ultimate goal is to build clinical decision support for fall prevention. Our model achieves higher precision and balanced accuracy than the commonly used multifactorial Missouri Alliance for Home Care fall risk assessment. This is the first known attempt to determine fall risk factors from the extensive OASIS data from a large sample. Our quantitative prediction of fall risks can aid clinical discussions of risk factors and prevention strategies for lowering fall incidence.

Keywords:

Falls, Health Risk Assessment, Machine Learning

Introduction

Falls are the leading cause of death due to injury in the home, especially for the elderly [1–3]. Falls are typically recurrent; those who fall once are two to three times more likely to fall again [4,5]. In 2015, direct costs related to fatal and non-fatal falls were $637.5 million and $31.3 billion respectively, making it one of the most costly patient safety problems among people aged 65 and older [6,7]. Identifying risk factors for falls is critical to the design of prevention protocols. Many research studies analyzed falls among hospitalized, long-term care, and community-dwelling older adults, yet few studies focus specifically on the frail, homebound population of older adults receiving home health care (HHC) services [8–10].

Medicare and agency policies direct clinicians to screen every HHC patient for fall risk, but only one HHC validated tool exists. The Missouri Alliance for Home Care assessment (MAHC-10) is a 10-item fall risk screening tool [11] with excellent sensitivity (97%), but with poor specificity (13%). In our study cohort, MAHC-10 identified over 93% of the cohort as having high fall risk, but only 5.14% actually had a fall (Figure 1). Therefore, using MAHC-10 as the default fall risk screening tool may increase the cost and burden of healthcare providers by triggering unnecessary provision of fall prevention strategies to almost every HHC patient. Moreover, MAHC-10 provides the clinician a score out of 10 instead of an actionable profile or fall risk as a probability per individual patient, making the design and implementation of personalized fall prevention difficult.

In this study, we devised a machine learning pipeline to explore the utility of existing HHC data containing rich patient information to predict fall risk. This is the first study to analyze large and comprehensive HHC datasets representing the characteristics of vulnerable older adults and the care they received in their homes to create models of effective fall risk prediction. With this larger dataset and additional input features from the electronic health record (EHR), we evaluated whether our models could achieve higher precision and accuracy than the existing risk scoring system.

Methods

Data Description

We leveraged patient information from the Centers for Medicare and Medicaid (CMS)-mandated Outcome and Assessment Information Set (OASIS) [12] data on an ethnically diverse population of nearly 60,000 patients from one large HHC agency in New York City. The OASIS is a mandatory detailed assessment with over 100 items evaluating a patient's clinical, behavioral, cognitive, and environmental conditions. The newer version (OASIS-C) was utilized for the patient population included in this study.

The HHC cohort was also assessed for fall risk using the 10-point Missouri Alliance for Home Care fall risk assessment (MAHC-10) [11]. A score greater than or equal to 4 is clinically regarded as at risk for falls. We took the intersection of patients who had both OASIS-C and MAHC-10 data and were over 65 years old, resulting in a final cohort of 59,028 unique patients. We supplemented the feature set with additional demographic information from the EHR, including language group and borough of residence (New York City).

Data Cleaning

We defined the binary outcome of fall incidence per patient from three sources: two OASIS-C items indicating whether the patient received emergent care or hospitalization due to falling (items M2310 and M2410), and whether a date of last fall was recorded in the EHR. Therefore, the outcome is True (presence of a fall) for a patient if the date of last fall is between the start and end of the HHC episode, or the answer is Yes for either M2310 or M2410. Using this definition, the fall incidence rate in our cohort is 5.14%. An episode of home care can be up to 60 days.

The OASIS-C start of care assessment contains 114 items, the majority being multiple-choice questions. Upon review from clinical experts, we considered 46 items that are relevant to fall risk prediction. For the multiple-choice OASIS-C items, we categorized the items based on the type of answer choices and applied the data cleaning strategy per category as follows.

For OASIS-C items with ordinal answer choices (for example, a 0-5 scale where 0 denotes never and 5 denotes all the time), we represented these items as continuous features. For OASIS-C items and EHR items with nominal answer choices (for example, gender, language group, type of assistance), we used one-hot encoding to create a binary feature per answer choice. The same encoding scheme applied for OASIS-C items that allowed multiple answers ("select all that applies"). We discarded patients under 65 years old to focus on the older adult population receiving HHC and used age as a continuous feature. Since OASIS-C is a mandatory assessment at the start of HHC, missing data is minimal (< 0.005%). We considered the rare missing entries as a separate answer choice for the corresponding questions.

Two OASIS-C items require listing diagnosis codes and severity of each diagnosis on a 4-point scale. These items were combined to reflect the overall physical well-being of a patient. To incorporate these items meaningfully into our feature set, we constructed summary features of diagnoses including: total number, total severity, and average severity. After data cleaning and feature engineering, our feature space contained 169 features from the OASIS-C and EHR that described the demographics, clinical, behavioral, and environmental characteristics of each HHC patient.

Machine Learning Pipeline

We devised a two-step machine learning pipeline with a feature selection step followed by a falls classification step. We randomly split the data into training (50%) and testing (50%) sets. Given the high-dimensionality and potential multi-collinearity of our feature set, we employed the ReliefF feature selection algorithm [13,14] to rank the features by the ReliefF score, computed based on the discriminative ability of each on the outcome, conditional on its neighboring features in the feature space.

We trained and compared random forest classifiers using three feature sets. First, we considered the OASIS-C features with positive ReliefF scores. We trained a random forest model with 5-fold cross-validation, using 300 estimators and default parameters as specified in the python scikit-learn [15] module *(OASIS model)*. Second, we took the ten items from MAHC-10 scoring as binary features and trained a random forest model using the same parameters as the OASIS model *(MAHC model)*. Finally, we explored if the features from OASIS-C and MAHC-10 together would augment the prediction accuracy of the random forest model. We combined the OASIS-C feature set with the 10 features from MAHC-10 to train a random forest classifier with the same parameter settings and 5-fold cross-validation *(Combined model)*.

Model Assessment

We evaluated the accuracy and precision of predictions from the OASIS model, the MAHC model, and the Combined model against the baseline MAHC-10 total score (hereafter called baseline), where a score greater than or equal to 4 is clinically considered at risk of falls. Since our outcomes are heavily skewed with only about 5% experiencing a fall, we used balanced accuracy as the metric to assess model accuracy. Balanced accuracy is defined as the average accuracy calculated from the two outcome classes. To highlight the performance of each model in correctly predicting the positive outcome, we contrasted the precision-recall curves of our models to that of the baseline. To summarize the precision-recall curves, we computed the average precision (AP), defined as the arithmetic mean of precisions at different recall thresholds.

Furthermore, we computed the area under ROC curves (AUC) for the test data corresponding to each feature set and contrasted them to the AUC of the baseline. To assess for significant differences between ROC curves across models, we performed bootstrapping on the test data to generate a confidence interval for each ROC curve. For each test set, we resampled with replacement for 1,000 times the outcome states and predicted probabilities, then calculated the AUC for each bootstrapped ROC curve. The 95% confidence interval of a ROC curve is given by all curves with AUC between the 2.5 and 97.5 percentile of all bootstrapped AUC values.

To evaluate the clinical significance of the OASIS model, we ranked the input features by their importance scores generated from the random forest classifier. To quantify the contribution of each feature to the accuracy of the model, we computed the gain in balance accuracy per added feature by refitting a random forest classifier iteratively, adding one feature at a time from the ranked feature list. Finally, by performing a hierarchical clustering of pairwise correlations between top-ranked OASIS-C features and the MAHC-10 items, we revealed fall risk factors identified by the OASIS model that were not captured in the baseline MAHC-10 assessment.

Results

We compared the random forest classifiers for three feature sets: 137 out of 169 OASIS- and EHR-derived features with positive ReliefF scores (OASIS model), MAHC-10 items (MAHC model), and OASIS-C plus MAHC-10 items (Combined model), in contrast to the MAHC-10 scoring (baseline). Overall, the MAHC model had comparable performance as the baseline, while the OASIS model and the Combined model had almost identical metrics that outperform the MAHC model and the baseline.

Although the baseline and the MAHC model both had an AUC of 0.6, the balanced accuracy of the MAHC model at 0.58 was slightly higher than that of baseline scoring measured at only 0.51. The OASIS model and Combined model both attained a balanced accuracy of 0.62 and an AUC of 0.67. The 95% bootstrap confidence interval of the OASIS model ROC was (0.66, 0.68), which was completely above the MAHC model ROC 95% confidence interval at (0.59, 0.62).

The precision of all models and the baseline was low, due to the low proportion of cases in our dataset. The average precision (AP) of the baseline was 0.07, and the AP of the MAHC model was 0.08. The OASIS and Combined models had an improved precision at AP=0.10. Consistent with the AP trend, the precision-recall curve of the OASIS model (and that of the Combined model) was above the curve of the MAHC model at all sensible recall thresholds (Figure 2).

Given the above metrics and the rule of parsimony, the OASIS model was the best out of the three models and the baseline. To investigate clinical relevance of the OASIS model, we ranked the input of 137 OASIS-C features by feature importance scores estimated by the classifier. The most important feature was age, with an importance score of 0.05, followed by the average and total severity of home care diagnoses. Frequencies of therapy visit and pain also had high feature importance scores. Balanced accuracy of the random forest classifier increased as features were added to the model in the order of importance; the balanced accuracy converged at around 0.62 after the top 45 features were added. The remaining 92 features had small contributions to the balanced accuracy of the random forest classifier.

To further evaluate the potential gain of using OASIS-C over MAHC-10 for fall risk prediction, we computed the pairwise correlation between each top-ranked OASIS-C feature among the MAHC-10 items, and performed a hierarchical clustering on the correlations. Four OASIS-C features had an analogous MAHC-10 item, including history of falls, visual impairment, cognitive impairment, and pain (Figure 3), as reflected by the strong positive correlations that were statistically significant (p-value < 10^{-5}). However, each of the four MAHC-10 items was also significantly correlated with a broad range of other OASIS-C items, indicating the heterogeneity among patients who scored the same on the MAHC-10 scale. In addition, some OASIS-C items were correlated in opposite directions to different MAHC-10 items, meaning that the effect of these features might be masked in the total MAHC-10 score. In particular, four top-ranked OASIS-C features were weakly correlated in opposite directions to MAHC-10 items but were not correlated to the total MAHC-10 score: frequency of ADL/IADL (activities of daily living/ instrumental activities of daily living) assistance, number of inpatient diagnoses, patient living alone, and patient living with others. These features provide new information on the patient's fall risk that was not available from the baseline scoring system.

Discussion

In patients receiving HHC services, falls rank as the top avoidable event that leads to disability, hospital admission and emergency department care [16,17]. Motivated by the low specificity of the existing fall risk assessment for HHC patients, we investigated the benefit of predicting fall risk using the Centers for Medicare and Medicaid Services (CMS) mandatory OASIS assessment for HHC coupled with supplemental EHR data. OASIS evaluates in detail the clinical, behavioral, cognitive, and environmental properties of a patient upon the start of a HHC episode. To analyze a large sample size of almost 60,000 and a high-dimensional input set comprising over 130 features with positive ReliefF score, we trained a random forest classifier on 50% of the sample and tested the accuracy and precision of the classifier on the remaining 50% of the data. Predictions leveraging this big data show improved accuracy and precision over a simplistic 10-point scoring system. We investigated if the ten items in the MAHC-10 would be more informative when used as features in a random forest model. We found a negligible change in precision and a slight improvement in balance accuracy in the MAHC model.

Furthermore, since all MAHC items are correlated to one or more OASIS items, complementing the OASIS features with the ten MAHC items did not result in detectable improvement over the OASIS-only model. This signals a potential to implement the OASIS model for clinical use, such that clinicians can obtain the per-patient fall risk from the model, once the mandatory OASIS assessment is completed. Clinicians burdened by required assessments and documentation may appreciate the efficiency gained.

Using our random forest classifiers, each patient in the test set gets an estimate of fall risk as a probability. This probability and the ranked list of important features are clinically relevant and potentially of great value when health care providers discuss fall risk with patients, allowing providers to customize and prioritize prevention strategies corresponding to actionable risk factors for each patient. Clinical decision support provided at the point of care may help clinicians target the most effective interventions and help patients recognize their risk and the impact of embracing preventive strategies to decrease their risk.

Limitations

This study used data from one large home care agency in New York City. The precision of the OASIS model is still lower than ideal for clinical use because the two outcome classes in our dataset are heavily imbalanced. Almost 95% of the patients in our cohort were not reported to have a fall. The low fall incidence from our structured data is consistent with the presumption that falls are often underreported. Recent reports in a national sample of older adults revealed 72% failed to report a fall when asked [18]. We hypothesize that a significant proportion of fall cases are recorded in the EHR narrative data instead of the structured data. Therefore, an immediate extension of our work is to use natural language processing techniques to identify additional fall instances from EHR narratives [19]. We also hope to extend the measurement period of falls beyond the home care episode. By having a more accurate estimate of fall incidence, we can fine-tune our classifiers and achieve higher precision.

Conclusions

This is the first known large-scale study to predict fall risk and characterize risk factors in the HHC community using the readily available OASIS assessment. By using machine learning models to analyze the rich feature set in a large cohort, we see promising improvement in the precision and accuracy of fall risk prediction over the MAHC-10 scoring system. Our results suggest that fall risk is a complex trait affected by a large number of risk factors of small effects. The machine learning approach also allows us to predict fall risk as a probability for each patient, and rank the importance of each risk factor. Our model confirms that a broad range of factors including age, clinical diagnoses, daily habits, living environment and hygiene, all contribute to a patient's fall risk. Further study incorporating an expanded feature set from the EHR will improve the estimate of fall incidence and fall risk prediction.

Figures and Graphs

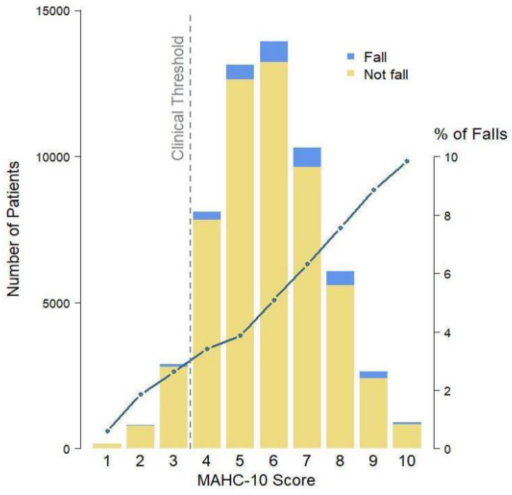

Figure 1. Number of patients per MAHC-10 score. Each bar shows the observed fall cases in blue and the non-cases in yellow. The black line shows the percentage of fall case per score. The clinical threshold of high fall risk is MAHC >=4 (Dashed vertical line).

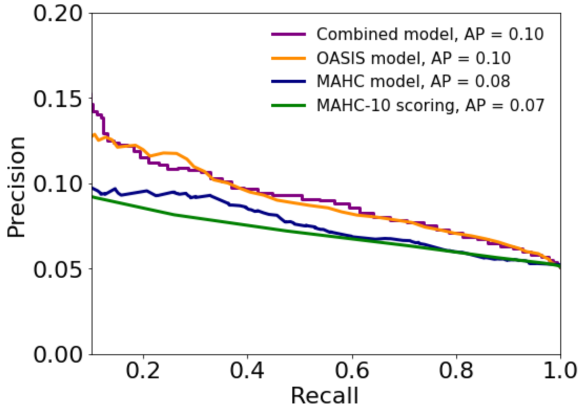

Figure 2. Precision-recall curves of the baseline scoring, MAHC model, OASIS model, and Combined model. For clarity, the panel only shows recall > 0.1 and precision < 0.2.

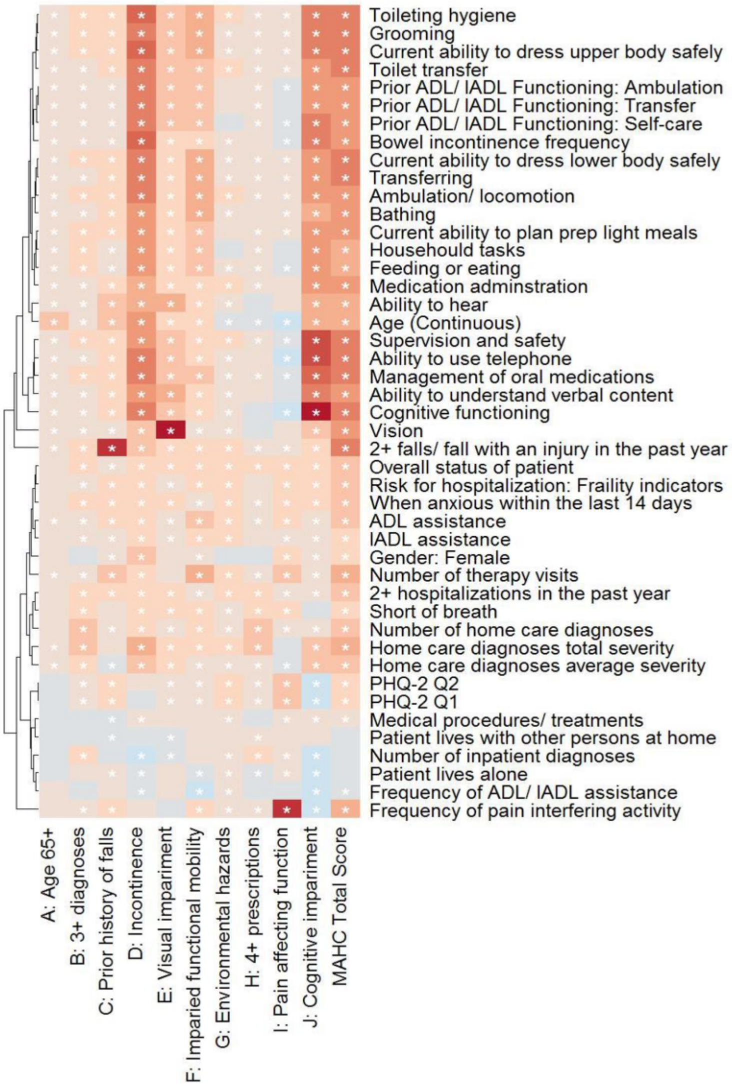

Figure 3. Hierarchical clustering of pairwise correlation (ρ) between top OASIS features (rows) and MAHC-10 items (columns). Blue represents negative correlation (-1 < ρ < 0) and red represents positive correlation (0 < ρ < 1). The asterisk (*) in a cell represents significant correlation after Bonferroni correction for multiple testing ($p < 10^{-5}$).

Acknowledgements

This study was supported by the University of Pennsylvania School of Nursing Frank Morgan Jones Fund. We thank Danielle Mowery, PhD for her valuable comments on the study methods.

References

[1] Centers for Disease Control and Prevention, Important Facts about Falls, (2017). https://www.cdc.gov/homeandrecreationalsafety/falls/adultfalls.html.
[2] K. Cameron, E. Schneider, D. Childress, C. Gilchrist C, Falls Free®: 2015 National Falls Prevention Action Plan, 2015. https://www.ncoa.org/wp-content/uploads/FallsActionPlan_2015-FINAL.pdf.
[3] G. Bergen, M.R. Stevens, and E.R. Burns, Falls and Fall Injuries Among Adults Aged ≥65 Years - United States, 2014, *MMWR Morb. Mortal. Wkly. Rep.* **65** (2016) 993–998.
[4] J.L. O'Loughlin, Y. Robitaille, J.F. Boivin, and S. Suissa, Incidence of and risk factors for falls and injurious falls among the community-dwelling elderly, *Am. J. Epidemiol.* **137** (1993) 342–354.
[5] World Health Organization, WHO global report on falls prevention in older age, (2007). http://www.who.int/ageing/publications/Falls_prevention7March.pdf.
[6] E.R. Burns, J.A. Stevens, and R. Lee, The direct costs of fatal and non-fatal falls among older adults - United States, *J. Safety Res.* **58** (2016) 99–103.
[7] Institute for Healthcare Improvement, No place like home: Advancing the safety of care in the home, (2018). http://www.ihi.org.
[8] C.L. Lewis, M. Moutoux, M. Slaughter, and S.P. Bailey, Characteristics of individuals who fell while receiving home health services, *Phys. Ther.* **84** (2004) 23–32.
[9] T. Yamashita, H. Jeon, A.J. Bailer, I.M. Nelson, and S. Mehdizadeh, Fall risk factors in community-dwelling elderly who receive Medicaid-supported home- and community-based care services, *J. Aging Health*. **23** (2011) 682–703.
[10] A.L. Byers, T. Sheeran, A.E. Mlodzianowski, B.S. Meyers, P. Nassisi, and M.L. Bruce, Depression and risk for adverse falls in older home health care patients, *Res. Gerontol. Nurs.* **1** (2008) 245–251.
[11] MAHC 10 Fall Risk Assessment Tool, Missouri Alliance for Home Care, n.d. https://www.homecaremissouri.org/projects/falls/documents/Oct2012FINALValidatedFallriskassessmenttool.pdf.
[12] Outcome and Assessment Information Set (OASIS), (2018). https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/OASIS/index.html.
[13] R.J. Urbanowicz, R.S. Olson, P. Schmitt, M. Meeker, and J.H. Moore, Benchmarking Relief-Based Feature Selection Methods for Bioinformatics Data Mining, *J. Biomed. Inform.* (2018). doi:10.1016/j.jbi.2018.07.015.
[14] I. Kononenko, Estimating attributes: Analysis and extensions of RELIEF, in: Machine Learning: ECML-94, Springer Berlin Heidelberg, 1994: pp. 171–182.
[15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, Scikit-learn: Machine Learning in Python, *J. Mach. Learn. Res.* **12** (2011) 2825–2830.
[16] Centers for Medicare and Medicaid Services, OASIS-C Based Home Health Agency Patient Outcome, Process and Potentially Avoidable Event Reports, (Updated June 2018). https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/OASIS/09aa_hhareports.html.
[17] M.D. Courtney, H.E. Edwards, A.M. Chang, A.W. Parker, K. Finlayson, and K. Hamilton, A randomised controlled trial to prevent hospital readmissions and loss of functional ability in high risk older adults: a study protocol, *BMC Health Serv. Res.* **11** (2011) 202.
[18] G.J. Hoffman, J. Ha, N.B. Alexander, K.M. Langa, M. Tinetti, and L.C. Min, Underreporting of Fall Injuries of Older Adults: Implications for Wellness Visit Fall Risk Screening, *J. Am. Geriatr. Soc.* **66** (2018) 1195–1200.
[19] V.J. Zhu, T.D. Walker, R.W. Warren, P.B. Jenny, S. Meystre, and L.A. Lenert, Identifying Falls Risk Screenings Not Documented with Administrative Codes Using Natural Language Processing, *AMIA Annu. Symp. Proc.* **2017** (2017) 1923–1930.

Address for Correspondence

Corresponding author: Yancy Lo, yancylo@pennmedicine.upenn.edu