A simplified approach to identification of risk status in patients with atherosclerotic cardiovascular disease

Aparna Sajja, MD a,*, Hsin-Fang Li b, Kateri J. Spinelli b, Amir Ali c, Salim S. Virani d, Seth S. Martin a, b, **, Ty J. Gluckman b, **

a Cincinnat i Center for the Prevention of Cardiovascular Disease, Division of Cardiology, Department of Medicine, Johns Hopkins University School of Medicine, 1800 Orleans Street, Baltimore, MD 21287, United States
b Center for Cardiovascular Analytics, Research and Data Science (CARDS), Providence Heart Institute, Providence Research Network, Portland, OR, United States
c Evaluation and Research, Providence Research Network, Renton, WA, United States
d Section of Cardiology, Michael E. DeBakey Veterans Affairs Medical Center and Section of Cardiovascular Research, Baylor College of Medicine, Houston, TX, United States

ARTICLE INFO

Keywords:
Cholesterol
Lipid
ASCVD
Secondary prevention

ABSTRACT

Objective: The 2018 American Heart Association/American College of Cardiology (AHA/ACC) Blood Cholesterol Guideline recommendation to classify patients with atherosclerotic cardiovascular disease (ASCVD) as very high-risk (VHR) vs not-VHR (NVHR) has important implications for escalation of medical therapy. We aimed to define the prevalence and clinical characteristics of these two groups within a large multi-state healthcare system and develop a simpler means to assist clinicians in identifying VHR patients using classification and regression tree (CART) analysis.

Methods: We performed a retrospective analysis of all patients in a 28-hospital US healthcare system in 2018. ICD-10 codes were used to define the ASCVD population. Per the AHA/ACC Guideline, VHR status was defined by ≥2 major ASCVD events or 1 major ASCVD event and ≥2 high-risk conditions. CART analysis was performed on training and validation datasets. A random forest model was used to verify results.

Results: Of 180,669 ASCVD patients identified, 58% were VHR. Among patients with a history of myocardial infarction (MI) or recent acute coronary syndrome (ACS), 99% and 96% were classified as VHR, respectively. Both CART and random forest models identified recent ACS, ischemic stroke, hypertension, peripheral artery disease, history of MI, and age as the most important predictors of VHR status. Using five rules identified by CART analysis, fewer than 50% of risk factors were required to assign VHR status.

Conclusion: CART analysis helped to streamline the identification of VHR patients based on a limited number of rules and risk factors. This approach may help improve clinical decision making by simplifying ASCVD risk assessment at the point of care. Further validation is needed, however, in more diverse populations.

Introduction

The 2018 American Heart Association/American College of Cardiology (AHA/ACC) Blood Cholesterol Guideline recommends risk stratification of patients with clinical atherosclerotic cardiovascular disease (ASCVD) to identify those at very high-risk (VHR) for future events (1). VHR is defined as having a history of two or more major ASCVD events (recent acute coronary syndrome (ACS), history of myocardial infarction (MI), ischemic stroke or symptomatic peripheral artery disease (PAD)) or 1 major ASCVD event and two or more high-risk conditions (age ≥65 years, diabetes, hypertension, smoking, familial hypercholesterolemia, chronic kidney disease, congestive heart failure, persistently elevated low density lipoprotein cholesterol (LDL-C), or prior coronary artery revascularization). Because lower levels of LDL-C correlate with reduced burden of atherosclerosis and better outcomes in secondary prevention of cardiovascular disease (2), high (or maximal) intensity statin therapy is recommended for all ASCVD patients. Guidance is further provided about the role of non-statin therapy in those with a persistent LDL-C level ≥70 mg/dL; however, the strength of recommendation and the choice of therapies (e.g., ezetimibe, PCSK9 inhibitor) vary among those felt to be VHR vs not-VHR (NVHR).
Given the number of factors to consider when assessing VHR status, it may well go unassigned in busy hospital or clinic settings. For some clinicians, this may be related to challenges in easily accessing relevant data at the point of care. It is also not clear whether redundancy exists among the 13 factors identified. Machine learning (ML) appears to be a useful approach to help address this, having previously been shown to improve classification based on guideline recommendations (3-5). In particular, a tree-based method represents a useful framework for risk classification (6). Accordingly, we sought to a) define the clinical characteristics of VHR and NVHR patients within a large multi-state healthcare system and b) develop a simpler means for assigning patients to one of these two groups utilizing classification and regression tree (CART) analysis.

Methods

We performed a retrospective analysis of electronic health record (EHR) data from 28 hospitals in a large multi-state healthcare system in the western US. Patients aged ≥18 years with clinical ASCVD and at least one lipid panel with triglycerides <400 mg/dl between January 1, 2018 and December 31, 2018 were included. Clinical ASCVD was defined by the presence of one or more related ICD-10 codes (Supplemental Appendix) in the patient’s history or problem list. For patients identified with ASCVD in 2018, the history and problem list was back-checked through October 1, 2015 (the start of ICD-10 codes), to capture diagnostics codes and social history relevant to VHR status. Outpatient encounter data was the primary data source, with hospital data included for identification of high-risk conditions. If multiple lab values were available, only the most recent was used.

VHR status was defined by ≥2 major ASCVD events (ACS during 2018, history of MI prior to 2018, ischemic stroke at any time, or symptomatic PAD at any time) or 1 major ASCVD event and ≥2 high-risk conditions (age ≥65 years, diabetes, hypertension, smoking, familial hypercholesterolemia, chronic kidney disease, congestive heart failure, persistently elevated LDL-C, or prior coronary artery revascularization). For smoking status, data from the patient’s most recent social history was used. Chronic kidney disease was defined by an estimated glomerular filtration rate (eGFR) of 15–59 mL/min/1.73m². Persistently elevated LDL-C was defined by an LDL-C ≥ 100 mg/dL while on statin therapy and ezetimibe in 2018. Patients not meeting the above criteria for VHR status were classified as NVHR. This study was approved by the Providence St. Joseph Health institutional review board, with waiver of informed consent. Aggregate data used in this study are available from the corresponding author upon reasonable request.

Patient demographics and data related to the 13 factors used to define VHR status were compared for VHR and NVHR patients. Categorical variables were described using frequencies and percentages. Continuous variables were described using means and standard deviations (SD).

To better identify the clinical and demographic factors associated with VHR status, a CART analysis was performed with VHR status as the primary outcome. The model selects a) variables most greatly associated with the outcome of choice and b) the preferred splitting point among all values to best classify observations into groups. For each subgroup, one additional splitting variable and point were chosen for evaluation. This process branches out continuously to reveal potential interactions, ultimately summarizing combinations of factors into an easily visualized tree-like plot (7,8). In our CART analysis, patients were randomly assigned equally into a training and testing set. The model was developed using the training set, and the resulting tree structure was evaluated using the testing set. A confusion matrix of actual and predicted VHR status was used to calculate specificity, sensitivity, and the ratio of misclassified individuals over the testing dataset.

Since results of the CART analysis were based on building a single tree, we also employed a random forest method to ensure that the CART variables selected were also important by alternative means. A random forest is built using a collection of random sampled trees (9,10) and does not yield the same tree-structure seen with a CART analysis. Instead, it produces a variable importance (VI) index for each variable that is computed based on an increase in misclassification when a given variable is excluded from the model. The VI indices can be used to rank the importance of variables relative to the outcome. Variables considered in both models included age, sex, race, ethnicity, and each of the VHR criteria.

The primary outcome for both models was VHR classification. Model performance was evaluated using area under the curve (AUC) and misclassification rates.

The following variables had missing data: race (n = 14,481), ethnicity (n = 66), and sex (n = 1). CART and random forest model analyses were performed using “party” package in R (Version 3.6.1). All other analyses were conducted using SAS Enterprise Guide 7.1 (SAS Institute Inc., Cary, NC, USA).

Results

A total of 180,669 patients with ASCVD were identified, with 105,124 (58.2%) classified as VHR and 75,545 (41.8%) classified as NVHR. The mean age and sex for these two groups were 73.1 ± 11.9 years, 55% male and 70.1 ± 13.4 years, 54% male, respectively (Table 1). Of the 105,124 VHR patients, 44,374 (42.2%) had two or more major ASCVD events and 60,750 (57.8%) had one major ASCVD event plus ≥2 high-risk conditions.

Among VHR patients, 47% had a recent ACS, 38% had a history of MI, 42% had an ischemic stroke, and 27% had symptomatic PAD (Table 2). The most prevalent high-risk conditions for those in this group were hypertension (88%), age ≥65 years (78%), diabetes (39%), and current smoking (37%). Patients with a recent ACS or prior history of MI were classified as VHR 96% and 99% of the time, respectively.

Of note, a sizable percentage of patients classified as VHR had one major ASCVD event plus two (18.4%), three (19.7%), or four (12.8%) high-risk conditions (Table 3). In contrast, the overwhelming majority (87.2%) of those classified as NVHR had no major ASCVD event. In this latter group, however, it was not uncommon for one (11.4%), two (22.3%), or three (24.1%) high-risk conditions to be found.

Patients with one major ASCVD event were classified as NVHR if they had <2 high-risk conditions. Most commonly, this included patients with ischemic stroke and no high-risk conditions (24.2%); these patients, however, represented only 3.1% of the NVHR group overall (Table 4). Instead, this population was largely made up of patients with ASCVD, but no major ASCVD event. This usually included patients with a) atherosclerotic heart disease of the native coronary artery without angina (36.2%), b) transient ischemic attack (TIA) (12.6%), c) other forms of cerebrovascular disease (6.8%), or d) other forms of coronary artery disease with coronary revascularization (6.6%) (Table 5).

After assigning patients to the training (n = 90,334) and testing (n = 90,335) datasets, recent ACS, ischemic stroke, hypertension, PAD, history of MI and age were identified as the most important predictors.

### Table 1: Patient Demographics.

| Variable | Overall | NVHR | VHR |
|----------|---------|------|-----|
| Patients, n (%) | 180,669 | 75,545 (41.8) | 105,124 (58.2) |
| Age, years | 72 ± 13 | 70 ± 13 | 73 ± 12 |
| Age ≥65 | 134,391 (74) | 51,955 (69) | 82,436 (78) |
| Sex, male | 99,365 (55) | 41,158 (54) | 58,207 (55) |
| Race | | | |
| White | 152,925 (92) | 64,945 (93) | 87,980 (91) |
| Black / African American | | | |
| Asian | 4742 (3) | 1463 (2) | 3289 (3) |
| Other* | 3057 (2) | 1162 (2) | 1892 (2) |
| Hispanic/Latino | 10,226 (6) | 3558 (5) | 6668 (6) |

Data presented as n (%) of patients or mean ± SD.

NVHR = not very high-risk, VHR = very high-risk.

* Native Hawaiian/Pacific Islander and American Indian/Alaska Native.
of VHR status (Fig. 1). Using this approach, we identified five groups with high likelihood of VHR status in the training dataset: 1) patients with recent ACS (96% VHR), 2) patients without recent ACS but with ischemic stroke and hypertension (96% VHR), 3) patients without recent ACS or stroke but with PAD (90% VHR), 4) patients without recent ACS, ischemic stroke, or PAD but with prior MI (92% VHR), and 5) patients without recent ACS or hypertension but with ischemic stroke and age over 65 years (72% VHR). When we classified patients in the validation set as VHR using these five rules, the CART model was associated with a sensitivity, specificity, positive predictive value, negative predictive value and misclassification rate of 99.5%, 91.3%, 94.1%, 99.3% and 3.9%, respectively. Similar findings were observed with the random forest model, where the most important VI index predictors (from highest to lowest) were ischemic stroke, recent ACS, PAD, history of MI, hypertension, and age (Supplemental Figure 1). The AUC for the CART and random forest models were 0.949 and 0.968, respectively.

### Discussion

While risk assessment represents a critical step in primary prevention of ASCVD (11), it has received appreciably less attention in secondary prevention. In spite of efforts to match the intensity of LDL-C reduction to the baseline risk of an individual (12), key lipid quality measures in this population largely employ a “one size fits all” approach (13,14).
A significant attempt to move beyond this took place with the 2018 AHA/ACC Blood Cholesterol Guideline, which was recommended that patients with ASCVD be further classified as VHR vs NVHR.

Using criteria recommended in the 2018 Blood Cholesterol Guideline, more than half (58%) of adults with ASCVD in the current study were classified as VHR. This rate is higher than that previously noted in a 2014–2015 analysis from the Veteran Affairs healthcare system (43%) (15); however, this may be related to differences in the populations studied. In contrast, a very similar rate of VHR patients was observed in a 2016 analysis of a large commercial and Medicare health insurance database (MarketScan) (55%), which interestingly had a low rate of patients from the western US (6%) (16).

Efforts focused on identifying VHR patients to date have largely highlighted significant gaps in care. In the aforementioned Veterans Affairs study, only 35% of VHR patients were receiving high-intensity statin therapy, 2% were on ezetimibe, and 67% had an LDL-C ≥ 70 mg/dL (15). Similar findings were noted in the MarketScan database, where only 35% of VHR patients were receiving high-intensity statin therapy, 7% were on ezetimibe, and 67% had an LDL-C ≥ 70 mg/dL (16). While not unique to VHR patients, a 2019 analysis of 2.6 million all-comer ASCVD patients in the NCDR PINNACLE registry demonstrated that 53% had never received lipid lowering therapy and among those on statin therapy (of any intensity), 68% had a LDL-C ≥ 70 mg/dL (17). Recently, in a 2020 analysis performed by Colantionio et al., authors also reported marked under-utilization of lipid lowering therapies in patients with ASCVD, coronary heart disease, and cerebrovascular disease (18,19). In that study, approximately 50% of patients with coronary heart disease and only one third of patients with PAD were on a statin.

Acknowledging that a number of factors likely contribute to marked underutilization of appropriate LDL-C lowering therapy in those with ASCVD (20–24), efforts to simplify personalized risk assessment may still be key. To this end, we utilized a CART analysis to identify major drivers of VHR status, and in the process, created simple rules to help guide treatment decision-making. Use of this approach allowed us to assess many levels of interactions between variables, as well as the impact of independent variables (7,25). Moreover, it allowed identification of interactions to be automated. Importantly, it also obviated the need to create frequency tables that include all of the risk factors and possible combinations for VHR and NVHR patients alike. Finally, it allowed for accommodation of larger samples with missing data, without losing power (26,27).

We believe that this CART-based approach provides a simplified means to guide risk assessment in secondary prevention. This is not a trivial issue, given the underutilization of known risk-reducing interven-

---

### Table 5

NVHR Patients with 0 Major ASCVD Events: Top 10 Diagnoses.

| ASCVD category | N (% of those with 0 major events) | % of NVHR group |
|----------------|------------------------------------|-----------------|
| Other CAD      | 28,315 (43.0%)                     | 37.0%           |
| Atherosclerotic heart disease of native coronary artery without angina pectoris (I25.1) | 27,719 (97.9%) | 36.2%           |
| Chronic ischemic heart disease, unspecified (I25.9) | 804 (2.8%) | 1.1%           |
| Coronary atherosclerosis due to calcified coronary lesion (I25.84) | 527 (1.9%) | 0.7%           |
| Atherosclerosis of coronary artery bypass graft w/o angina (I25.810) | 468 (1.7%) | 0.6%           |
| Coronary atherosclerosis due to lipid rich plaque (I25.83) | 337 (1.2%) | 0.4%           |
| TIA            | 9654 (14.7%)                       | 12.6%           |
| Other cerebrovascular disease | 5166 (7.9%) | 6.8%           |
| Other CAD + coronary revascularization | 5036 (7.7%) | 6.6%           |
| Other CAD + stable angina | 3193 (4.9%) | 4.2%           |
| Other arterial revascularization | 2561 (3.9%) | 3.4%           |
| Stable angina | 2341 (3.6%) | 3.1%           |
| Other CAD + other cerebrovascular disease | 1904 (2.9%) | 2.5%           |
| Other CAD + TIA | 1388 (2.1%) | 1.8%           |
| Other CAD + coronary revascularization + stable angina | 1251 (1.9%) | 1.6%           |

Abbreviations as in Table 2.

* N = 65,829 patients who are NVHR + 0 major ASCVD events.
tions in those with ASCVD (28), the significant residual risk faced by patients with ASCVD treated with high-intensity statin therapy (29,30), and the role risk plays in determining the cost-effectiveness of non-statin therapies (31,32). Simply embedding these rules into an EHR system without further vetting is likely to fail short (33). Nonetheless, there is a need to simplify the decision-making process for busy clinicians at the point of care, where risk assessment in secondary prevention is frequently underutilized. Future work should explore design and implementation of an EHR- or application-based tool, where clinicians may be guided by characteristics outlined in Fig. 1 to identify VHR patients.

Even for those with ASCVD initiated on appropriate LDL-C lowering therapy, adherence remains suboptimal (20,34,35). This, in part, is likely related to how risk is communicated by clinicians and ultimately perceived by patients (36,37). Prior studies have reported that those at VHR have approximately three times greater chance of developing future ASCVD events compared to those considered NVHR (16,38). Our results suggest that those classified as VHR using the five rules identified by our CART analysis leads to a low misclassification rate (3.9%), with less than half of the previously identified risk factors being considered. Of interest are the high-risk conditions that were not selected by the CART analysis (diabetes, smoking, familial hypercholesterolemia, chronic kidney disease, congestive heart failure, persistently elevated LDL-C, and previous coronary artery revascularization). While our study suggests these conditions do not need to be as strongly considered when evaluating ASCVD patients for VHR status, further validation is needed in more diverse populations, where the validity is currently unknown.

Our study has several limitations. First, patients in this study were disproportionately white and non-Hispanic. While race and ethnicity were not included in the AHA/ACC Blood Cholesterol Guideline as a means to define risk status, it remains unknown whether the same distribution of risk would have been observed in patients with greater racial and ethnic heterogeneity. Second, indices related to socioeconomic and insurance status, as well as medication (e.g., statin, ezetimibe, and PCSK9 inhibitor) utilization were not available. This data would have been particularly helpful in clarifying the magnitude and potential uppinings of care gaps. Third, while all patients in the dataset had their EHR queried, newer patients to the healthcare system may have had less complete documentation. Fourth, ICD-10 codes used to assess VHR status were derived from outpatient encounters and could not be tied directly to index hospitalizations. Fifth, while rules generated by the CART analysis provide an easy means to assess VHR predictors, it is limited by instability, such that small changes in partitioning can result in different tree structures. We attempted to overcome this, through inclusion of a random forest model which selected the same set of predictors. While the single tree developed from the CART analysis is supported by 500 trees from the random forest model, further validation is needed to confirm our model. Lastly, we did not assess the impact of the CART analysis on use of specific LDL-C lowering therapies. Accordingly, it will be important for future studies to examine the impact of this approach on treatment decisions at the point of care.

Conclusion

Limited data currently exist on the application of the 2018 AHA/ACC Blood Cholesterol Guideline in real-world settings. More than half of adults with ASCVD in our study met the definition of VHR, which is consistent with previous reports (15,16). In an attempt to simplify assignment of VHR status, however, we showed that use of five rules identified in our CART analysis could appreciably reduce the number of risk factors that need to be considered. Such an approach is appealing, in part, because EHR-based tools are either not widely available or underutilized. By considering a more limited number of risk factors, clinicians may be able to better employ the mental heuristics used to quickly tabulate VHR vs NHVR status. Such an approach simplifies ASCVD risk assessment and may help to streamline decision making at the point of care.

Author contributions

AS: Investigation, Writing – original draft and edits
HFL: Methodology, Formal analysis, Validation, Investigation, Writing – review and editing
KJS: Investigation, Project administration, Visualization, Writing – original draft and editing
AA: Data curation, Investigation, Software, Writing – review and editing
SSV: Investigation, Writing – original draft and edits
SSM: Investigation, Supervision, Writing – review and editing
TJG: Conceptualization, Methodology, Investigation, Supervision, Writing – original draft and editing

Declaration of Competing Interest

SSM: has current research support from the American Heart Association, PCORI, National Institutes of Health, the David and June Trone Family Foundation, and Pollin Digital Health Innovation Fund. He has served as a consultant to Astrazeneca, DalCor Pharmaceuticals, 89bio, Amgen, Esperion, Kaneka, and Sanofi. He has a patent System to estimate LDL cholesterol levels pending to Co-Inventor. He is a founder of and holds equity in Corrie Health, which intends to further develop the platform.

SSV: Grant support: Department of Veterans Affairs, World Heart Federation, Tahir and Jooma Family; Honorarium: American College of Cardiology (Associate Editor for Innovations, acc.org); Steering Committee Member: Patient and Provider Assessment of Lipid Management (PALM) registry [no financial remuneration].

AS, HFL, KJS, AA, TJG: None.

Sources of Funding

None.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ajpcrd.2021.100187.

References

[1] Grundy SM, Stone NJ, Bailey AL, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol. J Am Coll Cardiol 2019;73(24):e285–350.
[2] Bayoumy K, Gaber M, Mani P, et al. LDL-C targets in secondary prevention: how low should we go? Curr Cardiovasc Risk Rep 2019;13(8):23.
[3] Diller GP, Kempny A, Bahos-Naranjan SV, et al. Machine learning algorithms estimating prognosis and guiding therapy in adult congenital heart disease: data from a single tertiary centre including 10,019 patients. Eur Heart J 2019;40(13):1069–77.
[4] Kakadiaris IA, Vrigkas M, Yen AA, Kuznetsova T, Budoff M, Naghavi M. Machine learning outperforms ACC/AHA CVD risk calculator in MESA. J Am Heart Assoc 2018;7(22):e009476.
[5] Sardar P, Abbott JD, Kundu A, Aronow HD, Granada JF, Girij J. Impact of artificial intelligence on interventional cardiology. JACC Cardiovasc Interv 2019;12(14):1293–303.
[6] Banerjee M, Reynolds E, Andersson HB, Nallamothu BK. Tree-based analysis. Circ Cardiovasc Qual Outcomes 2019;12(5):e004879.
[7] Vyasnières MP, Plant RE, BH Allen-Diaz. Classification trees: an alternative non-parametric approach for predicting species distributions. J Veg Sci 2000;11(5):679–94.
[8] Nagy K, Reiczigel J, Hannos A, Schrott A, Kabi P. Tree-based methods as an alternative to logistic regression in revealing risk factors of crib biting in horses. J Equine Vet Sci 2010;30(1):21–6.
[9] Breiman L. Random forests. Mach Learn 2001;45(1):5–32 Oct1.
[10] Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application and characteristics of classification and regression trees, bagging and random forests. Psychol Methods 2009;14(4):233–48.
[11] Lloyd-Jones DM, Braun LT, Nudemle CE, et al. Use of risk assessment tools to guide decision-making in the primary prevention of atherosclerotic cardiovascular disease: a special report from the American heart association and American college of cardiology. J Am Coll Cardiol 2016;67(3):3254.
[12] Robinson JG, Huijgen R, Ray K, Persons J, Kastelein JJP, Pencina MJ. Determining when to add nonstatin therapy: a quantitative approach. J Am Coll Cardiol 2016;68(22):2412–21.
