Improved cardiovascular risk prediction using nonparametric regression and electronic health record data

Edward H. Kennedy, MS[1], **Wyndy L. Wiitala, PhD**[1], **Rodney A. Hayward, MD**[1,2], and **Jeremy B. Sussman, MD, MSc**[1,2]

[1]VA Center for Clinical Management Research, Ann Arbor VA Health Services Research and Development (HSR&D) Center of Excellence, Ann Arbor, MI

[2]Department of Internal Medicine and the Robert Wood Johnson Foundation Clinical Scholars Program, University of Michigan, Ann Arbor, MI

Abstract

Background—Use of the electronic health record (EHR) is expected to increase rapidly in the near future, yet little research exists on whether analyzing internal EHR data using flexible, adaptive statistical methods could improve clinical risk prediction. Extensive implementation of EHR in the Veterans Health Administration (VHA) provides an opportunity for exploration.

Objectives—To compare the performance of various approaches for predicting risk of cerebro- and cardiovascular (CCV) death, using traditional risk predictors versus more comprehensive EHR data.

Research Design—Retrospective cohort study. We identified all VHA patients without recent CCV events treated at twelve facilities from 2003 to 2007, and predicted risk using the Framingham risk score (FRS), logistic regression, generalized additive modeling, and gradient tree boosting.

Measures—The outcome was CCV-related death within five years. We assessed each method's predictive performance with the area under the ROC curve (AUC), the Hosmer-Lemeshow

Corresponding Author: Jeremy B. Sussman, 300 North Ingalls, Room 7C11, Ann Arbor, MI 48109. Phone: 734-936-4787. Fax: 734-936-8944. jeremysu@med.umich.edu.

SUPPLEMENTAL DIGITAL CONTENT
Supplemental Digital Content 1.
Table. Inclusion criteria for sample
Supplemental Digital Content 2.
Appendix A. EHR Data in the VA
Supplemental Digital Content 3.
Appendix B. Nonparametric Regression
Supplemental Digital Content 4.
Figure. Observed / expected CCV events by estimated risk
Supplemental Digital Content 5.
Figure. Estimates of CCV risk across prediction methods

Disclaimers: The authors have no conflicts to report. The opinions expressed here are those of the authors and do not necessarily represent those of the Department of Veterans Affairs.

goodness-of-fit test, plots of estimated risk, and reclassification tables, using cross-validation to penalize over-fitting.

Results—Regression methods outperformed the FRS, even with the same predictors (AUC increased from 71% to 73% and calibration also improved). Even better performance was attained in models using additional EHR-derived predictor variables (AUC increased to 78% and net reclassification improvement was as large as 0.29). Nonparametric regression further improved calibration and discrimination compared to logistic regression.

Conclusions—Despite the EHR lacking some risk factors and its imperfect data quality, healthcare systems may be able to substantially improve risk prediction for their patients by using internally-developed EHR-derived models and flexible statistical methodology.

INTRODUCTION

Heart attack and stroke are the first and fourth leading causes of death in the United States and together cost nearly $300 billion annually.[1] Further, cardiac risk is a major predictor of treatment benefit[2] and plays a central role in treatment guidelines for cardiovascular medication use[3,4] and diagnostic testing.[5] Thus, accurate prediction of cerebro- and cardiovascular (CCV) risk in individual patients is a major clinical and research focus, bringing with it critical implications for physician decision-making and for research on comparative and cost effectiveness.[6] Risk prediction is no simple task, however, particularly since it is dependent on both the quality and amount of available data as well as on statistical methods used for estimation. Although there is a great deal of research in the development of new tests to help assess cardiovascular risk,[7–9] less attention has been directed to whether the increased availability of large amounts of clinical data in the EHR can allow for more accurate risk assessment tools. The effectiveness of current tools is limited by the populations, datasets, and statistical tools that enabled their development.

Current risk prediction tools in America are based on cohort studies that use chart review and patient interviews to obtain high-quality data. They require personnel time to calculate, are based on particular patient populations with relatively limited sample sizes, and are difficult and expensive to update.[10] When used in external populations years after their initial development, the validity of these tools can be inconsistent.[10] However, an increasingly available electronic health record (EHR) may provide an opportunity to make internally developed risk tools that are more accurate and timely.[11,12] Further, the use of EHRs is expected to increase rapidly with a recent multi-billion-dollar investment by the federal government.[13]

Thus far, integrated and comprehensive EHR systems have seen limited use in clinical practice.[14] However, the Veterans Health Administration (VHA) is an exception,[15] and provide a unique opportunity for exploring the benefits and opportunities of EHR.

Clinical risk prediction also could benefit from concurrent progress in the development of flexible and adaptive (or "nonparametric") regression and machine learning methods, many of which are particularly well suited for large datasets with many variables.[16] For example, "ensemble methods" (which are also known as "meta-classifiers", and include random

forests and boosting) work by incorporating predictions from a large number of small, simple models, and would have been computationally infeasible until relatively recently. These and related methods provide "automatic" prediction since they do not require the user to specify whether or how to include predictor variables, are robust to outliers, and limit overfitting by design.[16]

In this study, we used EHR data across multiple VHA hospitals with flexible statistical methods to explore the extent to which risk prediction of fatal CCV events can be improved relative to traditional approaches. Specifically, we compared predictive performance using: (1) the Framingham Risk Score (FRS) versus internally-developed EHR-derived models; (2) parametric versus nonparametric regression methodology; and (3) traditional risk predictors versus additional risk predictors obtainable from the VHA EHR (such as hypertensive medications and comorbid conditions).

METHODS

Data and sample

We used data from patients treated at twelve Veterans Health Administration (VHA) facilities in the midwestern United States from 2003 to 2007. The outcome of interest was CCV-related death within five years, and was determined using ICD-10 codes I00–I99 from the National Death Index (NDI). Unlike non-fatal CCV events, this outcome is available even for patients who discontinue care at VHA facilities.

Our eligibility criteria were based on those used by the Office of Quality & Performance (OQP) for the VA External Peer Review Program (EPRP). We included in our sample all patients over 18 years of age with at least two visits to a primary care, cardiology/ hypertension, endocrine/diabetes, chronic infectious disease care or mental health clinic during the baseline year (2003). The visit criterion was used to exclude patients who attend VHA facilities infrequently for prescription medication refills, while using outside facilities for clinical care. In order to ensure that patients did not have a recent history of cardiovascular disease, we excluded those with a known CCV diagnosis or event during the baseline year, using information from the VHA Medical SAS Data Sets. This makes the sample more comparable to that of the Framingham study. The table in Supplemental Digital Content 1 shows the sample size for each inclusion criterion.

We obtained information on diagnoses from the VHA Medical SAS Data Sets, medication and laboratory data from the VHA Decision Support System files (DSS), and vital signs from the VHA Corporate Data Warehouse (CDW). For patients who had more than one recorded lab value, we used the average of the last two during the baseline year, and we considered biologically implausible values to be missing. We used ICD-9 codes to identify baseline status for clinical diagnoses that are also candidate predictors of CCV risk (diabetes: 250.xx, 362.0x, 357.2x, 366.41; chronic obstructive pulmonary disease (COPD): 490.x, 491.x, 496.x; inflammatory arthritides: 710.x, 714.x, 720.x, 711.1, 711.3, 713.1; periodontal disease: 523.4, 523.9; sleep apnea: 327.23).[17–20] Diabetes status was also determined using pharmacy information (any insulin or oral hypoglycemic prescription fills). All risk predictor data was recorded during the baseline year, and missing values were

singly imputed via chained equations.[21] More details about our data and about EHR data in the VHA in general are available in Supplemental Digital Content 2, Appendix A.

Prediction methods

For our primary analyses, we compared four risk prediction methods: the Framingham risk score (FRS, which used Weibull regression), parametric logistic regression, nonparametric generalized additive modeling (GAM), and boosting (a nonparametric ensemble method) across three sets of predictors. See Figure 1 for a schematic summarizing the different types of prediction methods; we chose one representative method from each type. These four methods are all well established but technically very different.

Since mortality was the outcome of interest, we implemented the FRS that predicts CCV death.[22] This FRS is derived from a Weibull survival model including sex, age, systolic blood pressure, smoking status, total/HDL cholesterol ratio, diabetes, and ECG-left ventricular hypertrophy (ECG-LVH). Smoking status and ECG-LVH were unavailable for our sample. The model assumed by FRS for the probability of CCV-related death within five years (denoted by Y) is:

$$P(Y{=}1|X{=}x){=}1{-}\exp\left\{-\exp\left\{\frac{\log(5)-\mu(x)}{\exp[\theta+\theta_1\mu(x)]}\right\}\right\}, \text{with}\,\mu(x){=}\beta_0{+}\sum_{j=1}^{p}\beta_j x_j$$

where $x = (x_1, ..., x_p)$ is the observed vector of predictors, j indexes the p elements of x, and $\beta = (\beta_0, \beta_1, …, \beta_p)$, θ, and θ_1 are parameters with specified values estimated from the Framingham cohort.

To tailor the FRS to our sample, we also estimated a "recalibrated" FRS that used the transformed FRS as a predictor in a logistic regression model:

$$\text{logit}\,P(Y{=}1|X{=}x){=}\alpha_0{+}\alpha_1\left\{\frac{\log(5)-\mu(x)}{\exp[\theta+\theta_1\mu(x)]}\right\}$$

with α_0 and α_1 parameters estimated in our sample, and $\mu(x)$ as defined previously.

Logistic regression is by far the most popular method for general prediction of a binary outcome; it generally relies on exact specification of a parametric model.[16,23] Our assumed logistic regression model was:

$$\text{logit}\,P(Y{=}1|X{=}x){=}\gamma_0{+}\sum_{j=1}^{p}\gamma_j x_j$$

where the parameter vector $\gamma = (\gamma_0, \gamma_1, …, \gamma_p)$ was estimated in our sample. We log-transformed systolic blood pressure, medication counts, number of NEXIS visits during the baseline year, and serum creatinine, and otherwise assumed a linear relationship between predictors and log-odds of risk. We did not consider interaction terms, and no variable

selection criteria were applied so all predictors were included. Logistic regression can be viewed as yielding a model similar to the FRS Weibull model, but with all coefficients tailored to our sample.

The generalized additive model (GAM)[24] is similar to standard logistic regression, except nonparametric functions f_j of the predictors are estimated (via splines) instead of coefficients for linear terms.[16,24] The assumed model is:

$$\text{logit}P(Y=1|X=x)=f_0+\sum_{j=1}^{p}f_j(x_j)$$

where f_0 is an intercept term. We implemented GAM using the "gam" package in R, using default settings. As with our implementation of logistic regression, no variable selection criteria were applied and all predictors were included in each model.

Boosting[25] is an "ensemble method" (or "meta-classifier") that combines many simple models (e.g., trees with very few splits) fit to successively re-weighted or re-sampled data using weights that increase the importance of previously misclassified data points.[16,26,27] Unlike previously discussed methods, boosting automatically performs variable selection and does not yield a simple regression formula that can be written out. We implemented gradient tree boosting using the "gbm" package in R with 500 two-level trees, thereby allowing all possible pairwise interactions.

More details and further explanation about GAM and boosting are available in Supplemental Digital Content 3, Appendix B. We also used multivariate adaptive regression splines (MARS)[28] and random forests[29] in secondary analyses. Results are available upon request.

Risk predictors

We explored predictive performance across three subsets of risk predictors: (1) only the traditional risk predictors used to calculate the FRS, (2) the traditional risk predictors along with medication information, and (3) traditional risk predictors with medication information, lab values, vital signs, diagnoses, and other data (see Table 1 for a list of all candidate predictors).

Evaluation

We assessed discrimination and calibration of the four risk prediction methods using the area under the ROC curve (AUC) and the Hosmer-Lemeshow goodness-of-fit (HL-GOF) test, respectively, with 10-fold cross-validation to penalize overfitting.[16,30] We explored changes in estimated risk with a reclassification table and the net reclassification improvement (NRI)[31] using (0–2%, 2–5%, 5–10%, 10–20%, 20–100%) risk groupings, and also present plots of the distributions of estimated risk.

All statistical analyses were conducted using R[32] and Stata.[33]

RESULTS

Patient characteristics

The characteristics of the patients in the sample are given in Table 1. There were 113,973 patients total and 4,995 (4.4%) died within five years of CCV-related causes. Most patients (95.6%) were male and the mean age was 63.4 years (SD = 13.6 years). The rate of comorbidity was high, including high rates of diabetes (24.3%), COPD (24.1%), hypertension treatment (68.9% prescribed at least one medication at baseline), and narcotic or opiate prescriptions (25.6%). Patients who died of CCV-related causes were on average 10.1 years older at baseline, 45% more likely to have a diagnosis of diabetes, and more likely to be prescribed medications (26% more likely for hypertension and 47% more likely for diabetes).

Discrimination and calibration

Table 2 shows cross-validated discrimination and calibration results. All estimates were calculated using conservative cross-validation methodology (with k=10 folds) to penalize overfitting. Of note, the use of more predictor variables yielded larger increases in discrimination than the use of more flexible analytic methods. Using only traditional risk predictors, AUC increased from 71.3% for the FRS to 72.6% for logistic regression and to 73.1% for GAM and boosting. With more predictors the advantage of using nonparametric as opposed to logistic regression was magnified. For example, when using all selected predictors rather than just the traditional set, AUC increased by 3.7 points for logistic regression, 4.4 points for GAM, and 4.7 points for boosting. Discrimination and calibration results for MARS and random forests were similar to those for GAM and boosting (available upon request).

The FRS predicted only 3,031 CCV-related deaths, compared to the 4,995 that were actually observed (39.3% fewer). Calibration for the FRS (p<0.001 for HL-GOF test) was particularly poor for low-risk patients; the FRS predicted 89.2% fewer deaths than observed for patients in the lowest quintile of risk. However, there was no evidence of poor calibration for the recalibrated FRS (HL-GOF p=0.236). Boosting also predicted fewer deaths than observed for low-risk patients, but only when using all available data (p=0.017 for HL-GOF test). The figure in Supplemental Digital Content 4 shows plots of observed versus expected numbers of events when using all predictors, as in traditional Hosmer-Lemeshow tables. Using a 5% risk threshold with all available predictors, cross-validated sensitivity was 42.4% for the FRS, 68.3% for logistic regression, 67.0% for GAM, and 63.6% for boosting; cross-validated specificity was 82.3% for the FRS, 71.1% for logistic regression, 73.8% for GAM, and 77.3% for boosting.

To assess whether missing EHR data on smoking history could fully account for the under-prediction of CCV events in the FRS, we performed a sensitivity analysis. Specifically, we examined the most favorable scenario for FRS by assuming every patient with a CCV-related death was a smoker and by randomly assigning a positive smoking status to 30% of the rest of the population. In this scenario, the number of CCV events predicted by FRS only

increased to 3659 (26.8% fewer than observed), and the HL-GOF test still rejected at the 0.05 level (p<0.001).

Reclassification and distribution of estimated risk

To further examine the predictiveness and calibration of the models, we conducted reclassification analyses based on risk groups of 0–2%, 2–5%, 5–10%, 10–20%, and 20100% (Table 3). Internally-developed models always yielded significantly improved net reclassification compared to the externally-developed FRS, and nonparametric methods gave significant improvement over logistic regression.

The largest net reclassification improvement (NRI) was for boosting relative to the FRS (NRI: 0.29, 95% CI: 0.27–0.32). Compared to the FRS, boosting reclassified 61.7% of those with an event to a higher risk group and 8.3% of those without an event to a lower risk group. The smallest NRI was for boosting relative to GAM (NRI: 0.07, 95% CI: 0.06–0.09); boosting only reclassified 15.6% of those with events to a higher risk group and 10.5% of those without events to a lower risk group, compared to GAM. NRI was not significant comparing the FRS with the recalibrated FRS (NRI: 0.00, 95% CI: −0.02 to 0.02), indicating that recalibration did not achieve significantly better arrangement into the specified risk categories despite improved calibration.

Plots of estimated risk show the underestimation of the FRS for low-risk patients, along with the relative similarity among internally-developed methods; these plots are shown in Supplemental Digital Content 5.

These findings do have clinical implications. For example, consider what would happen if aspirin were used for anyone with an estimated risk greater than 5%, and we could estimate risk using either the FRS or boosting. Using boosting, 12,568 people would be treated who would not have been treated under the FRS, and 5,767 people would not be treated who would have been treated under the FRS. The boosting treatment regime would be more accurate, since 1,366 of the 12,568 people (10.9%) who would have been treated only when using boosting went on to die of CCV-related causes, compared to only 203 of the 5,767 (3.5%) who would have been treated only when using the FRS.

Relationships between predictors and risk

Figure 2 shows the relationship between traditional risk predictors (i.e., those used to calculate the FRS) and estimated risk of 5-year CCV-related death for reference patients whose other predictor values were set equal to their respective sample means (see Table 1 for specific values). We found that less flexible methods mischaracterized the association between some predictors and the risk of CCV-related death. Most notably, the FRS assumes strong increasing relationships for systolic blood pressure (SBP) and cholesterol ratio, but in our sample these relationships were quite weak. Further, GAM and boosting showed evidence of a slight U-shaped risk profile for SBP (i.e., patients with very low or very high SBP were at highest risk), but this was missed by logistic regression since it assumed linearity. Thus, not only would a different coefficient value for SBP in the FRS be appropriate for our data, but the SBP variable itself would require additional transformation to maximize model fit.

DISCUSSION

Risk prediction is recommended in the management of every class of clinical cardiovascular decision, including diagnostic testing, aspirin, cholesterol, and anti-hypertensive use, and has a 50-year history in cardiology practice. In spite of this, the most commonly-used risk scores are developed in selected populations, with limited sample sizes and risk predictors, using rigid statistical tools. Using the VHA as an example, we have shown that some of these limitations can be overcome with use of the EHR and nonparametric regression techniques. Flexible risk models that are developed internally or "recalibrated" can greatly improve predictive performance in terms of both discrimination and calibration. The best way to achieve greater performance seems to be by including more risk predictors, even when using imperfect EHR data. Using nonparametric methods with traditional risk predictors gave modest improvements compared to parametric approaches, but larger improvements occurred when additional predictors were added. Differences between the nonparametric methods (GAM and boosting) were not definitive. Even greater improvements could become more apparent with larger datasets (allowing for consideration of more variables, interactions, and transformations) and with the automation of these prediction tools. Such automation would obviate the need for data entry, and could facilitate rapid implementation and result in dramatic benefits to patient care. Our results did not have the benefit of large, expensive chart review or survey data but were comparable with those of larger studies.[34,35]

Our findings should be interpreted in the context of a number of limitations. First, this is a feasibility study and the models we developed require additional work before consideration for use in clinical practice. Second, although the quality of EHR data in the VHA is impressive relative to many other data sources, it is still imperfect. For example, some variables frequently had missing values, while others were missing completely (such as smoking status and ECG-LVH). The EHR design could make these data elements searchable and text searching techniques are rapidly improving. The exclusion of smoking status and ECG-LVH could negatively impact the FRS in particular, though data enhancements should presumably improve the performance of all methods and sensitivity analyses suggested that increased performance of the FRS would be modest. Third, although we excluded patients with a recent CCV event, we did not have data to exclude all patients with any history of cardiovascular disease, which would make our sample more like the one in which the FRS was developed. Next, while our new data and methods improved prediction substantially, AUC was never more than 80%; some patients' risk will inevitably still be misclassified. Finally, although fatal CCV events are an important outcome, non-fatal cardiovascular events are more commonly assessed in the risk prediction literature. We chose this outcome partly because it is reliably available in VA data, unlike non-fatal CCV events.

This study also has a number of important strengths. First, we used a large dataset (over 100,000 patients) that is representative of a meaningful and sizable portion of the VHA population. Second, we used a reliable outcome derived from the National Death Index (NDI), which is not always the case for EHR research. Finally, we have meaningfully extended the literature on risk prediction by assessing both state-of-the-art methods and varying amounts of data.

To our knowledge, investigation of the dual impact of data quality and statistical methodology is rare. Those studies that have appropriately assessed the use of nonparametric methods did not focus on advantages of EHR,[36,37] and a recent review[36] indicates that others typically made limited comparisons (e.g., only between logistic regression and classification trees) or had methodological flaws (e.g., no cross-validation). Although Wu et al.[38] recently provided an excellent comparison of nonparametric methods in a machine learning context while also discussing EHR issues, they did not explicitly evaluate the benefit of using comprehensive EHR data versus only traditional risk predictors. Also, of the two nonparametric approaches they explored, one (support vector machines) gave poor performance and the other (AdaBoost) has an improved implementation.[16]

Some of our findings were particularly surprising. For example, the U-shaped and weakened associations between blood pressure and cholesterol and CCV-related death are in strong distinction to most risk scores. Recognizing that our goal is prediction and not causal inference, we believe these associations could be due to the reported phenomenon of treatment lowering LDL and SBP more than CCV mortality.[39,40]

This study suggests a number of opportunities for future work, both with respect to methodology and implementation in clinical practice. Extending this work to larger databases, and adding changes over time and increased clinical data (such as diagnostic study results) could allow for substantially greater predictive accuracy. An exploration of non-fatal CCV events would be useful to determine whether the impact of flexible methods and more data on quality of prediction would change. Further, simulation-based cost-effectiveness studies could draw on our results to fully assess the practical benefit and feasibility of better risk prediction. Our study shows that, once the data exists, developing good risk scores is within the scope of most large health systems. In fact, the nonparametric methods explored here are no more difficult to implement than logistic regression and are accessible in freely available software. The greatest opportunity for future work, though, is in clinical and policy implementation. The drive towards patient-centered, tailored care[41] will make risk prediction and personalization more important than ever.

The recent political incentives towards Accountable Care Organizations[42] and EHR[43] will lead to a radical increase in large clinical datasets that make within-population clinical risk prediction viable for many healthcare systems. Our results suggest that these healthcare systems could achieve better risk prediction by using internally-developed models with EHR data and flexible statistical methodology.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

REFERENCES

1. Roger VL, Go AS, Lloyd-Jones DM, et al. Heart disease and stroke statistics-2012 update: a report from the American Heart Association. Circulation. 2012; 125(1):e2–e220. [PubMed: 22179539]

2. Hayward RA, Krumholz HM, Zulman DM, et al. Optimizing statin treatment for primary prevention of coronary artery disease. Ann Intern Med. 2010; 152(2):69–77. [PubMed: 20083825]

3. Aspirin for the prevention of cardiovascular disease: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2009; 150(6):396–404. [PubMed: 19293072]

4. Chobanian AV, Bakris GL, Black HR, et al. Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. Hypertension. 2003; 42(6):1206–1252. [PubMed: 14656957]

5. Greenland P, Bonow RO, Brundage BH, et al. ACCF/AHA 2007 clinical expert consensus document on coronary artery calcium scoring by computed tomography in global cardiovascular risk assessment and in evaluation of patients with chest pain: a report of the American College of Cardiology Foundation Clinical Expert Consensus Task Force (ACCF/AHA Writing Committee to Update the 2000 Expert Consensus Document on Electron Beam Computed Tomography). Circulation. 2007; 115(3):402–426. [PubMed: 17220398]

6. Lloyd-Jones DM. Cardiovascular risk prediction: basic concepts, current status, and future directions. Circulation. 2010; 121(15):1768–1777. [PubMed: 20404268]

7. Lee TH, Boucher CA. Clinical practice. Noninvasive tests in patients with stable coronary artery disease. N Engl J Med. 2001; 344(24):1840–1845. [PubMed: 11407346]

8. Greenland P, LaBree L, Azen SP, et al. Coronary artery calcium score combined with Framingham score for risk prediction in asymptomatic individuals. JAMA. 2004; 291(2):210–215. [PubMed: 14722147]

9. Ridker PM, Buring JE, Rifai N, et al. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. JAMA. 2007; 297(6): 611–619. [PubMed: 17299196]

10. Cooney MT, Dudina AL, Graham IM. Value and limitations of existing scores for the assessment of cardiovascular risk: a review for clinicians. J Am Coll Cardiol. 2009; 54(14):1209–1227. [PubMed: 19778661]

11. Chaudhry B, Wang J, Wu S, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med. 2006; 144(10):742–752. [PubMed: 16702590]

12. Blumenthal D, Glaser JP. Information technology comes to medicine. N Engl J Med. 2007; 356(24):2527–2534. [PubMed: 17568035]

13. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med. 2010; 363(6):501–504. [PubMed: 20647183]

14. Jha AK, DesRoches CM, Campbell EG, et al. Use of electronic health records in U.S. hospitals. N Engl J Med. 2009; 360(16):1628–1638. [PubMed: 19321858]

15. Jha AK, Perlin JB, Kizer KW, et al. Effect of the transformation of the Veterans Affairs Health Care System on the quality of care. N Engl J Med. 2003; 348(22):2218–2227. [PubMed: 12773650]

16. Hastie, T.; Tibshirani, R.; Friedman, JH. The elements of statistical learning : data mining, inference, and prediction. 2nd ed.. New York, NY: Springer; 2009.

17. Sin DD, Man SF. Chronic obstructive pulmonary disease as a risk factor for cardiovascular morbidity and mortality. Proc Am Thorac Soc. 2005; 2(1):8–11. [PubMed: 16113462]

18. Banerjee S, Compton AP, Hooker RS, et al. Cardiovascular outcomes in male veterans with rheumatoid arthritis. Am J Cardiol. 2008; 101(8):1201–1205. [PubMed: 18394459]

19. Deliargyris EN, Madianos PN, Kadoma W, et al. Periodontal disease in patients with acute myocardial infarction: prevalence and contribution to elevated C-reactive protein levels. Am Heart J. 2004; 147(6):1005–1009. [PubMed: 15199348]

20. Yaggi HK, Concato J, Kernan WN, et al. Obstructive sleep apnea as a risk factor for stroke and death. N Engl J Med. 2005; 353(19):2034–2041. [PubMed: 16282178]

21. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Stat Med. 1999; 18(6):681–694. [PubMed: 10204197]

22. Anderson KM, Odell PM, Wilson PW, et al. Cardiovascular disease risk profiles. Am Heart J. 1991; 121(1 Pt 2):293–298. [PubMed: 1985385]

23. Hosmer, DW.; Lemeshow, S. Applied logistic regression. 2nd ed.. New York: Wiley; 2000.

24. Hastie, T.; Tibshirani, R. Generalized additive models. 1st ed.. London ; New York: Chapman and Hall; 1990.

25. Freund, Y.; Schapire, R. Proceedings of the Thirteenth International Conference on Machine Learning. San Francisco: Morgan Kaufmann; 1996. Experiments with a new boosting algorithm.

26. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat. 2001; 29(5): 1189–1232.

27. Buhlmann P, Hothorn T. Boosting algorithms: regularization, prediction and model fitting. Stat Sci. 2007; 22(4):477–505.

28. Friedman JH. Multivariate adaptive regression splines. Ann Stat. 1991; 19(1):1–67.

29. Breiman L. Random forests. Mach Learn. 2001; 45(1):5–32.

30. Harrell, FE. Regression modeling strategies : with applications to linear models, logistic regression, and survival analysis. New York: Springer; 2001.

31. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med. 2008; 27(2):157–172. [PubMed: 17569110]

32. R: a language and environment for statistical computing [computer program]. Vienna, Austria: R Foundation for Statistical Computing; 2008.

33. Stata Statistical Software: Release 12 [computer program]. College Station, TX: StataCorp LP; 2011.

34. Thomsen TF, McGee D, Davidsen M, et al. A cross-validation of risk-scores for coronary heart disease mortality based on data from the Glostrup Population Studies and Framingham Heart Study. Int J Epidemiol. 2002; 31(4):817–822. [PubMed: 12177028]

35. Conroy RM, Pyorala K, Fitzgerald AP, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. Eur Heart J. 2003; 24(11):987–1003. [PubMed: 12788299]

36. Austin PC. A comparison of regression trees, logistic regression, generalized additive models, and multivariate adaptive regression splines for predicting AMI mortality. Stat Med. 2007; 26(15): 2937–2957. [PubMed: 17186501]

37. Austin PC, Tu JV, Lee DS. Logistic regression had superior performance compared with regression trees for predicting in-hospital mortality in patients hospitalized with heart failure. J Clin Epidemiol. 2010; 63(10):1145–1155. [PubMed: 20304609]

38. Wu J, Roy J, Stewart WF. Prediction modeling using EHR data: challenges, strategies, and a comparison of machine learning approaches. Med Care. 2010; 48(6 Suppl):S106–S113. [PubMed: 20473190]

39. Law MR, Wald NJ, Rudnicka AR. Quantifying effect of statins on low density lipoprotein cholesterol, ischaemic heart disease, and stroke: systematic review and meta-analysis. BMJ. 2003; 326(7404):1423. [PubMed: 12829554]

40. Law MR, Morris JK, Wald NJ. Use of blood pressure lowering drugs in the prevention of cardiovascular disease: meta-analysis of 147 randomised trials in the context of expectations from prospective epidemiological studies. BMJ. 2009; 338:b1665. [PubMed: 19454737]

41. Krumholz HM. Patient-centered medicine: the next phase in health care. Circ Cardiovasc Qual Outcomes. 2011; 4(4):374–375. [PubMed: 21772000]

42. Shortell SM, Casalino LP. Implementing qualifications criteria and technical assistance for accountable care organizations. JAMA. 2010; 303(17):1747–1748. [PubMed: 20442391]

43. Blumenthal D. Launching HITECH. N Engl J Med. 2010; 362(5):382–385. [PubMed: 20042745]
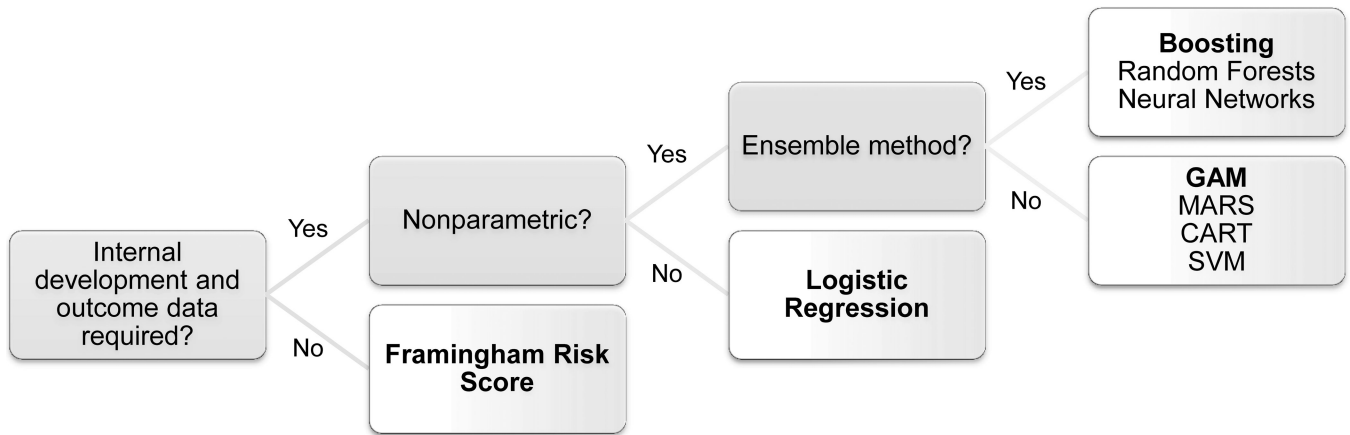
Figure 1. Schematic of prediction methods
Nonparametric methods do not require exact specification of a model for the outcome. Ensemble methods incorporate predictions from many small and simple models applied to, for example, weighted or sampled versions of the data. Bolded methods are those used in this paper. [Abbreviations: Generalized additive model (GAM), multivariate adaptive regression splines (MARS), classification and regression tree (CART), support vector machine (SVM).]

Age
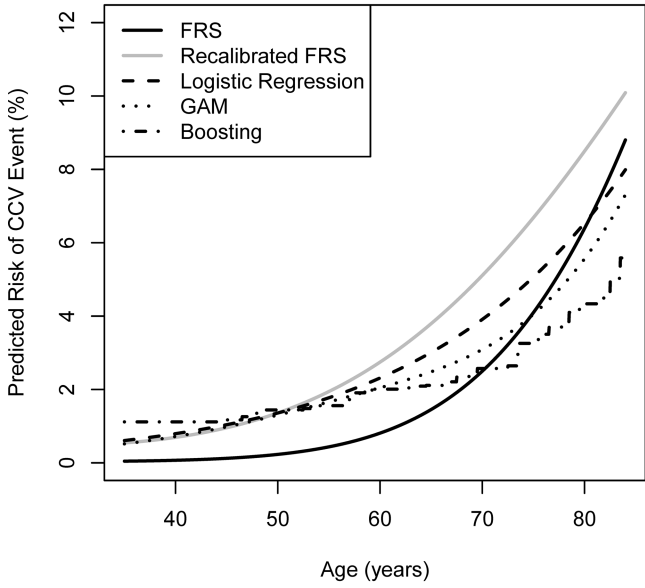
Systolic Blood Pressure
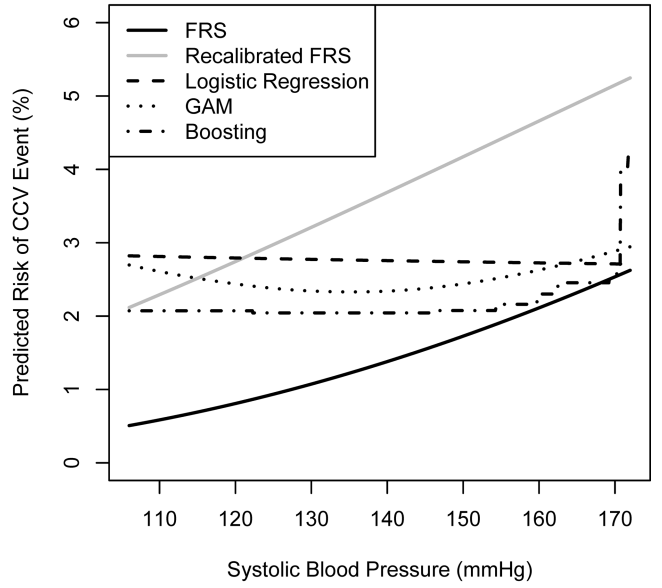
Total/HDL Cholesterol
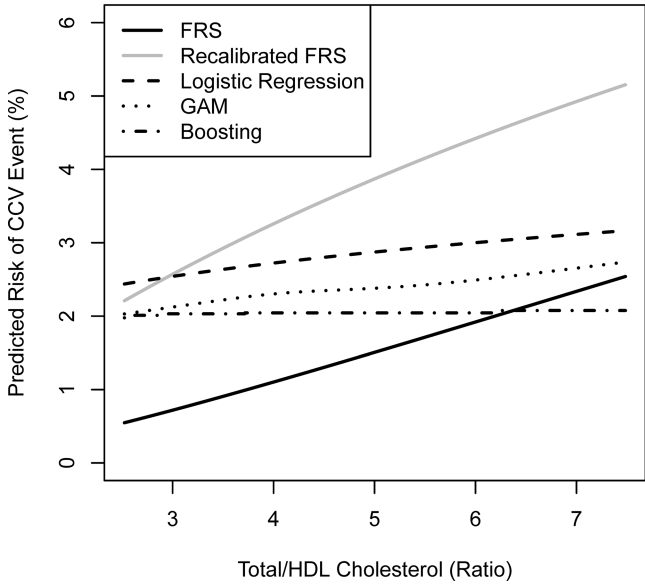
Diabetes
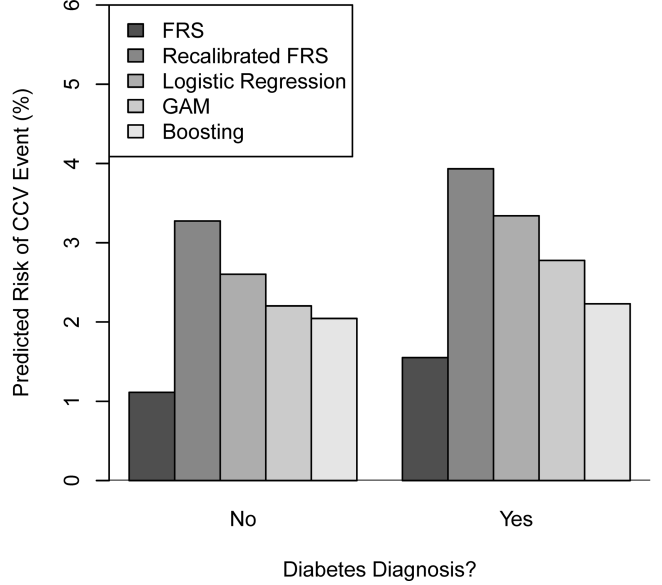
Figure 2. Estimated relationships between traditional predictors and CCV risk
Predictions were generated using only the traditional risk predictors (age, systolic blood pressure, cholesterol ratio, and diabetes status); risk was estimated across the middle 95% of observed values of the predictors of interest at the sample mean of all other predictors.

Table 1

Patient characteristics[a,b]

		CCV-related death?		
	All patients	**Yes**	**No**	**p-value**
Number of patients	113,973 (100%)	4,995 (4.4%)	108,978 (95.6%)	---
Traditional predictors				
Age	63.4 ± 13.6	73.0 ± 10.8	62.9 ± 13.5	< 0.001
Male	109,007 (95.6%)	4,927 (98.6%)	104,080 (95.5%)	< 0.001
Systolic BP	135.5 ± 16.3	138.5 ± 19.0	135.4 ± 16.1	< 0.001
Total/HDL chol. ratio	4.4 ± 1.2	4.3 ± 1.2	4.4 ± 1.2	< 0.001
Diabetes diagnosis	27,684 (24.3%)	1,726 (34.6%)	25,958 (23.8%)	< 0.001
Medication data [c]				
Hypertension meds	78,578 (68.9%)	4,308 (86.2%)	74,270 (68.2%)	< 0.001
Lipids	55,184 (48.4%)	2,649 (53.0%)	52,535 (48.2%)	< 0.001
Diabetes meds	24,940 (21.9%)	1,573 (31.5%)	23,367 (21.4%)	< 0.001
Narcotics or opiates	29,234 (25.6%)	1,315 (26.3%)	27,919 (25.6%)	0.270
Benzodiazepines	14,375 (12.6%)	678 (13.5%)	13,700 (12.6%)	0.052
Levothyroxines	7,989 (7.0%)	516 (10.3%)	7,473 (6.9%)	< 0.001
Anticoagulants	5,058 (4.4%)	580 (11.6%)	4,478 (4.1%)	< 0.001
Labs, vitals, diagnoses & other				
Albumin		3.9 ± 0.3	4.0 ± 0.3	< 0.001
Blood urea nitrogen	17.4 ± 6.5	21.6 ± 9.1	17.2 ± 6.3	< 0.001
LDL cholesterol	112.1 ± 27.5	104.8 ± 26.3	112.5 ± 27.5	< 0.001
Serum creatinine	1.1 (0.9–1.2)	1.2 (1.0–1.4)	1.1 (0.9–1.2)	< 0.001
Pulse	74.3 ± 12.3	74.1 ± 12.8	74.4 ± 12.3	0.180
Pulse pressure	60.5 ± 14.7	66.4 ± 16.8	60.3 ± 14.5	< 0.001
Baseline diagnoses:				
COPD	27,523 (24.1%)	1,455 (29.1%)	26,068 (23.9%)	< 0.001
Periodontitis	4,749 (4.2%)	93 (1.9%)	4,656 (4.3%)	< 0.001
Inflamm. arthritis	1,655 (1.5%)	76 (1.5%)	1,579 (1.4%)	0.720
Sleep apnea	1,174 (1.0%)	18 (0.4%)	1,156 (1.1%)	< 0.001
Body mass index	29.3 ± 5.6	28.2 ± 6.0	29.3 ± 5.6	< 0.001
Num. NEXIS visits	3 (2–5)	3 (2–5)	3 (2–5)	0.012

[a] For categorical variables, n (%) is shown; for continuous variables, mean ± sd is shown except in case of a skewed distribution, for which median (IQR) is shown.

[b] Among predictors with any missing values, missingness ranged between 2.8% (Systolic BP) and 45.7% (LDL cholesterol). Most patients (63.6%) had zero or one missing risk predictor value.

[c] Lipids include: statins, fibrates, ezetimibe. Narcotics/opiates include: buprenorphine, codeine, fentanyl, hydrocodone, hydromorphone, methadone, morphine, opium, oxycodone, pentazocine, propoxyphene, tramadol. Benzodiazepines include: alprazolam, chlordiazepoxide, clonazepam, clorazepate, diazepam, flurazepam, lorazepam, oxazepam, temazepam, triazolam. Anticoagulants include: dalteparin, enzoxaparin, fondaparinux, heparin, warfarin

Table 2

Cross-validated discrimination and calibration

Model	Traditional Risk Predictors			Traditional plus Medication			Traditional plus Medication plus Labs/vitals/diagnoses/other		
	AUC %, avg ±SE	HL-GOF X²±SE	p-val	AUC %, avg±SE	HL-GOF X²±SE	p-val	AUC %, avg±SE	HL-GOF X²±SE	p-val
FRS	71.3±1.0	479±122	<0.001	---	---	---	---	---	---
Recalibrated FRS	71.3±1.0	14.0±6.7	0.236	---	---	---	---	---	---
Logistic regression	72.6±1.0	18.8±6.6	0.064	74.3±0.8	13.0±3.9	0.175	76.3±1.0	9.1±3.8	0.404
GAM	73.1±0.9	10.6±3.0	0.273	74.8±0.7	10.3±4.4	0.338	77.5±0.9	11.1±3.5	0.250
Boosting	73.1±0.9	11.6±4.2	0.243	74.9±0.7	13.5±6.2	0.244	77.8±0.9	21.4±4.4	0.017

Abbreviations: area under the ROC curve (AUC), Hosmer-Lemeshow goodness of fit test (HL-GOF); Framingham risk score (FRS), generalized additive model (GAM)

Table 3

Reclassification using (0–2%, 2–5%, 5–10%, 10–20%, 20–100%) groupings[a,b,c]

	Logistic regression		GAM		Boosting	
	Lower Risk	Higher Risk	Lower Risk	Higher Risk	Lower Risk	Higher Risk
FRS						
No event	5035 (4.6%)	38220 (35.1%)	7024 (6.4%)	37023 (34.0%)	9087 (8.3%)	36813 (33.8%)
Event	293 (6.8%)	2886 (57.8%)	293 (5.9%)	2992 (59.9%)	349 (7.0%)	3081 (61.7%)
NRI (95% CI)	0.21 (0.19, 0.24)		0.27 (0.24, 0.29)		0.29 (0.27, 0.32)	
Logistic regression						
No event	---	---	10663 (9.8%)	8097 (7.4%)	17446 (16.0%)	11717 (10.8%)
Event	---	---	435 (8.7%)	742 (14.9%)	736 (14.7%)	1207 (24.2%)
NRI (95% CI)	---		0.09 (0.07, 0.10)		0.15 (0.13, 0.16)	
GAM						
No event	---	---	---	---	11393 (10.5%)	8279 (7.6%)
Event	---	---	---	---	572 (11.5%)	781 (15.6%)
NRI (95% CI)	---		---		0.07 (0.06, 0.09)	

Abbreviations: generalized additive model (GAM); net reclassification improvement (NRI), confidence interval (CI)

[a] Percentages in rows labeled "event" are among patients with CCV events; in rows labeled "no event" percentages are among patients without CCV events.

[b] Numbers in the "Lower Risk" columns reflect patients reclassified into lower risk groupings (e.g., 20–100% to 10–20%), while numbers in the "Higher Risk" columns reflect patients reclassified into higher risk groupings (e.g., 0–2% to 2–5%).

[c] For these results, the FRS uses only traditional predictors while other methods use all available predictors.