Original Report: Population Health Approaches to Reduce Chronic Disease Disparities

Evaluating Social Determinants of Health Variables in Advanced Analytic and Artificial Intelligence Models for Cardiovascular Disease Risk and Outcomes: A Targeted Review

Jane L. Snowdon, PhD[1]; Elisabeth L. Scheufele, MD, MS[1]; Jill Pritts, MS[1]; Phuong-Tu Le, BS, CEP[2]; George A. Mensah, MD[3]; Xinzhi Zhang, MD, PhD[3]; Irene Dankwa-Mullan, MD, MPH[1]

Introduction/Purpose: Predictive models incorporating relevant clinical and social features can provide meaningful insights into complex interrelated mechanisms of cardiovascular disease (CVD) risk and progression and the influence of environmental exposures on adverse outcomes. The purpose of this targeted review (2018–2019) was to examine the extent to which present-day advanced analytics, artificial intelligence, and machine learning models include relevant variables to address potential biases that inform care, treatment, resource allocation, and management of patients with CVD.

Methods: PubMed literature was searched using the prespecified inclusion and exclusion criteria to identify and critically evaluate primary studies published in English that reported on predictive models for CVD, associated risks, progression, and outcomes in the general adult population in North America. Studies were then assessed for inclusion of relevant social variables in the model construction. Two independent reviewers screened articles for eligibility. Primary and secondary independent reviewers extracted information from each full-text article for analysis. Disagreements were resolved with a third reviewer and iterative screening rounds to establish consensus. Cohen's kappa was used to determine interrater reliability.

Results: The review yielded 533 unique records where 35 met the inclusion criteria. Studies used advanced statistical and machine learning methods to predict CVD risk (10, 29%), mortality (19, 54%), survival (7, 20%), complication (10, 29%), disease progression (6, 17%), functional outcomes (4, 11%), and disposition (2, 6%). Most studies incorporated age (34, 97%), sex (34, 97%), comorbid conditions (32, 91%), and behavioral risk factor (28, 80%) variables. Race or ethnicity (23, 66%) and social variables, such as education (3, 9%) were less frequently observed.

Conclusions: Predictive models should adjust for race and social predictor variables, where relevant, to improve model accuracy and to inform more equitable interventions and decision making. *Ethn Dis.* 2023;33(1):33–43; doi:10.18865/ed.33.1.33

Keywords: Artificial Intelligence; Analytics; Predictive Models; Cardiovascular Disease; Health Disparities; Social Determinants of Health

[1] Center for Artificial Intelligence, Research, and Evaluation, IBM Watson Health, Cambridge, MA 02142
[2] Division of Integrative Biological and Behavioral Sciences, National Institute on Minority Health and Health Disparities, National Institutes of Health, Bethesda, MD 20892
[3] Center for Translation Research and Implementation Science, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892

Address correspondence to Jane L. Snowdon, PhD; IBM Corporation, Yorktown Heights, NY 10598. snowdonj@us.ibm.com

Introduction

Advanced analytics and artificial intelligence (AI) are powerful leading-edge methodologies for uncovering trends, hidden patterns, and insights to augment informed decisions in health care. They can be useful in enhancing our understanding of disease risk, progression, outcomes, and their correlation with health determinants. Over the past 2 decades, AI and machine learning (ML) techniques have been applied to inform a broad range of health care decisions and management processes across different clinical and public health domains. Some of these applications include promoting precision health and patient-centered care, optimizing clinical operational performance, enhancing treatment guidelines, and predicting disease risk and prognosis. These methodologies require data sources with considerable volume, variety, completeness, and timeliness to provide the accuracy and applicability needed for unbiased analyses.

The goal of this article is to examine the extent to which present-day advanced analytics, AI, and ML methodologies are addressing potential biases in the analytic and algorithmic models that inform care and management of patients with cardiovascular disease (CVD). Predictive models incorporating relevant clinical and social factors can provide meaningful insights into the complex interrelated mechanisms of CVD progression[1] and risk[2] as well as the influence of environmental exposures on adverse outcomes. In addition to augmenting clinical decision and care management, these models can also generate

predictions about future CVD-related morbidity and mortality and inform tailored behavioral and lifestyle interventions to prevent adverse events.

Cardiovascular disease remains the leading cause of mortality in the United States and the world despite many advances in understanding disease development, risk, and treatment. It is well known that comorbidities, such as mental health status, type 2 diabetes mellitus (T2DM), obesity, and being overweight or physically inactive, can adversely affect risk and outcomes.[3] Research provides evidence that factors such as age, race or ethnicity, income, sex, food security, and even community location are also associated with disease development and outcome. For example, the third National Health and Nutrition Examination Survey (NHANES III) has provided longitudinal insights into many factors that affect the risk and development of CVD. One study determined that the likelihood of developing subclinical myocardial injury was less likely for Mexican Americans (odds ratio [OR], 0.74; 95% confidence interval [CI], 0.62-0.88) compared with Whites or Blacks and that subclinical myocardial injury increased the risk of cardiovascular mortality (hazard ratio [HR], 1.74; 95% CI, 1.13-2.67; $P<.05$).[4] Other factors, such as income, have also been associated with CVD risk. In a 10-year study of NHANES data and predicted absolute cardiovascular risk, the percentage of adults at risk for CVD by 20% or more declined with time but not for those with incomes at or below the federal poverty level.[5]

In addition to the effect on outcomes for individuals with coexisting chronic conditions and social disadvantage, numerous noted disparities exist in the published literature on the management and treatment of CVD.[6,7] In a study of sex disparities in adults enrolled in commercial health plans, women with T2DM were less likely than men with T2DM to have their low-density-lipoprotein cholesterol controlled at <100 mg/dL. Race or ethnicity inequities continue to exist despite a long-standing awareness of these issues. A study of race or ethnicity disparities in relation to CVD and risk for individuals living in New York City found that non-Latino Black women were at higher risk of being overweight or obese, having hypertension, and having T2DM than non-Latino White men or women.[8]

Despite the growing body of knowledge relating social determinants of health (SDoH) to CVD risk and disease progression, limited evidence is available on the extent to which relevant factors are currently being incorporated into clinical predictive models to improve performance accuracy. Increasing evidence suggests that clinical risk and predictive models can be flawed due to inherent biases in the data sources, coding of the predictors, sample size, and/or study design as well as implicit biases in selected outcomes or in validating or applying a variety of advanced analytical and AI methodologies.[9–12] Data sources may not be fully representative of a population of interest, resulting in model underperformance or unreliable predictions that may perpetuate inequities.[13] In a published state-of-the-art review of precision health analytics and implementation research from the *Journal of the American College of Cardiology*, authors called for more actionable models that could translate across communities and subpopulations, including adjustment for relevant social risk factors.[14]

An abundance of predictive models have been published in the past 2 decades around cardiovascular disease, some of which have been incorporated into clinical practice guidelines.[15] A number of systematic reviews on cardiovascular disease risk modeling have also been published, testing validity of these models for incorporation into evidence-based practice and in various US population cohorts.[16]

This targeted review, using systematic methods, collected recently published PubMed literature (2018–2019) and sought to identify the extent to which race, ethnicity, and other SDoH factors were being used to develop clinical predictive models about CVD in North America. As knowledge of the effect of bias in data and prediction models advances, we aim to highlight potential challenges and propose to enrich data sets for modeling with more representative predictors.

Methods

A targeted review was conducted of the PubMed literature to search, identify, and critically evaluate primary studies published in English that reported on or described predictive models for cardiovascular disease, associated risks, and outcomes in the general adult population in North America. Additional scope included limiting the review to a 2-year time frame from January 1, 2018, to December 31, 2019, and to assess the extent to which the models presented in the literature incorporated sociodemographic information, social determinants of health, and relevant predictors of CVD.

A single reviewer conducted the search and retrieved the relevant articles using the prespecified inclusion and exclusion criteria given in Table 1. Excluded articles were those that focused on pediatric populations since our focus was on adults and on stroke and stroke management because brain and vascular imaging information were key factors for stroke management and formed the basis of clinical decision making. We excluded studies that focused on CVD imaging, as our focus was on cardiovascular risk. Predictive models for heart transplants were excluded, as those focused on popula-

Table 1. Inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
The publication reports on a predictive or descriptive model about cardiovascular disease with clearly defined risks and outcomes in humans.	The publication is not focused on a predictive or descriptive model about cardiovascular disease to address risks and outcomes in humans.
	The publication focuses on stroke or CVD imaging or heart transplant or congenital heart disease or reports only on a refinement or valuation of a prediction model.
The publication reports on an adult population in the United States or Canada.	The publication reports on a pediatric population, or the publication reports on an adult population outside of North America.
The publication reports on a primary study or systematic review with or without meta-analysis.	The publication reports a narrative (nonsystematic) review, case report, case series, editorial, perspective, commentary, or single abstract without a full paper for reference.
The publication date is within the 2 years 2018–2019.	The publication date is not within the 2 years 2018–2019.
The publication is in English.	The publication is in a language other than English.

tions with congenital disease. Articles that reported only on a refinement or valuation of the accuracy of a prediction model or algorithm were also excluded, as were editorials, perspectives, commentaries, single abstracts without full papers for reference, and studies conducted outside of the United States and Canada.

Two independent reviewers were assigned to each article to screen for eligibility based on the title, abstract, and setting of the study. After inclusion consensus was reached about the article, primary and secondary independent reviewers were assigned to each full-text article to extract information for analysis. The primary reviewer independently reviewed the full paper and categorized the type of methodology, AI or ML approach, CVD outcomes, and range of predictors included in the model. The secondary reviewer cross-checked and validated the categorization. Disagreements were resolved with a third reviewer and iterative screening rounds to establish consensus. For consistency in data extraction and coding the predictors, a standardized Excel data extraction form was used. Quantitative synthesis was conducted of the predictors or factors used, AI or ML and advanced analytics approaches used, and CVD outcomes used for the clinical prediction models. Cohen's kappa was used to determine interrater reliability.

The analysis focused on published articles of models that described CVD care in the areas of risk stratification, prediction of disease outcomes, and CVD treatment and were also considered to be of clinical significance. Articles that used univariate or multivariable regression analysis, ML, and natural language processing approaches for estimating the risk of CVD, risk of complications, risk of mortality, survival outcome, disease progression, and functional outcome were included. The models were examined to identify which social determinant predictor variables were incorporated according to the categories of sex, race or ethnicity, education level, employment status, disability, comorbid conditions, activity limitations, region or environmental exposures, rurality vs urbanicity, behavioral risk factors such as smoking, and type of insurance. An assessment was made to determine whether the model considered 1 or more of the social determinant variables as well as the effect of the features on CVD risk, progression, or outcomes.

Results

The rapid review literature search strategy identified 533 potentially relevant articles after duplicate removal. Article titles and abstracts reviewed for eligibility excluded 375 papers based on the exclusion criteria (Table 1). Further screening of 158 potentially relevant, eligible articles eliminated an additional 116 full-text papers (ie, the study focused on different outcomes than CVD, such as stroke, or the predictive factor was related to heart transplant and management). After further screening, 42 papers were identified for full-text screening. In total, 35 articles met criteria for inclusion[17–51] (Figure 1). Interrater reliability of eligibility determination had an overall weighted kappa of 91% for full-text screening.

Details of results are available in supplemental tables from the corresponding author (Supplementary Tables 1–3). All studies were conducted in North America, primarily in the United States,[17–34,36–51] with 3 in Canada.[34,35,49]

Figure 2 shows that most studies used advanced statistical models, such as logistic[*] (20, 57%) and Cox (also known as proportional hazards) regression[†] (13, 37%). A few studies used ML[24,46,51] (3, 9%), linear regression[51] (1, 3%), and multivariable adjusted competing-risks regression[38]

[*] *References 18-20, 22-25, 27, 29, 31, 32, 34, 35, 37, 41, 44-46, 48, 51.*

[†] *References 17, 18, 21, 29, 33, 36, 38, 39, 42, 43, 45, 49, 50.*

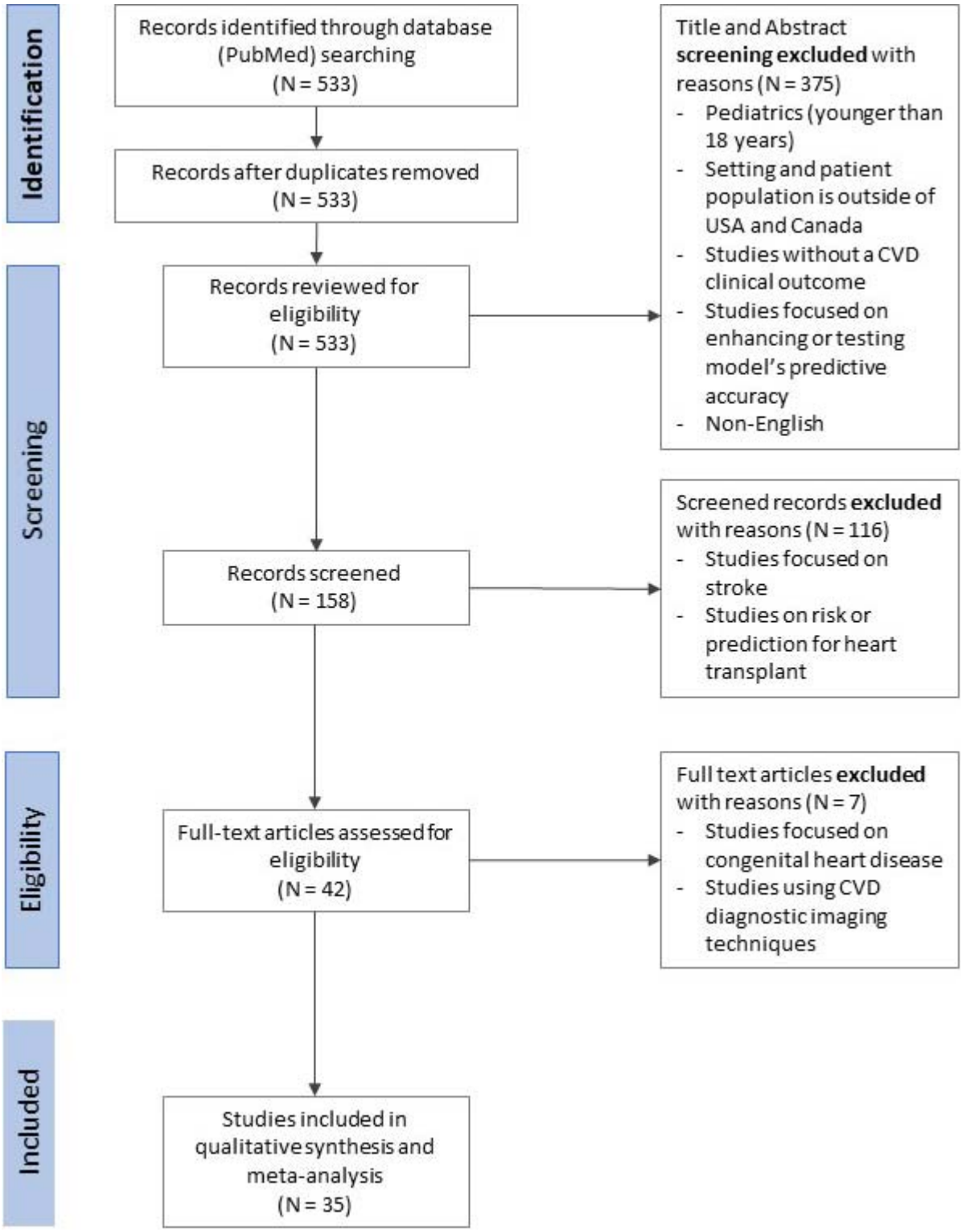

Figure 1. Article disposition and literature screening flow diagram for the rapid review phases

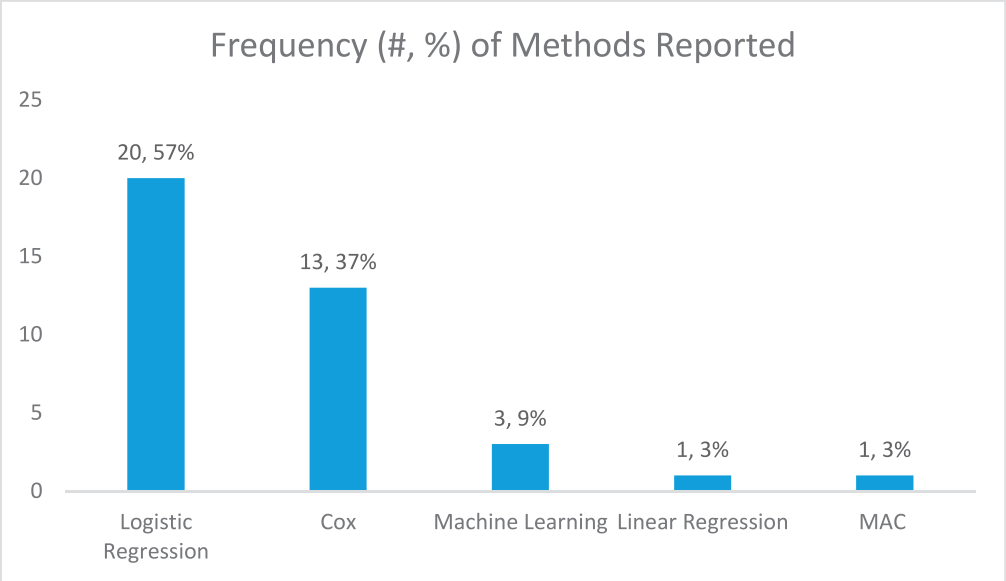

Figure 2. Frequency of advanced analytic and AI methodologies reported

(1, 3%). In many cases, studies used a combination of models.

Figure 3 summarizes the range of predictive and descriptive models that were applied to data sets to analyze and predict CVD risk or outcomes. Models focusing on risk stratification investigated outcome of long-term survival[17,21,23,27,42,45,46] (7, 20%), outcome of discharge disposition[21,25] (2, 6%), and risk of functional outcome[19,22,23,31] (4, 11%). Models focusing on prediction of disease outcomes examined risk of developing a CVD event[18,26,29,33,36,38,39,43,48,50] (10, 29%), risk of CVD complications[19–21,30–32,34,40,41,45] (10, 29%), and risk of disease progression[22,33,36,44,48,49] (6, 17%). More than half of the studies reported on risk of mortality posttreatment or postintervention[‡] (19, 54%).

In terms of the variables or features selected for inclusion in the models, most studies incorporated sex[17–46,48–51] (34, 97%), presence of comorbid conditions[17–40,42,43,45,47–51] (32, 91%), and behavioral risk factors[§] (eg, physical activity, diet, or smoking) (28, 80%), while few studies examined activity limitations[18,23,25,30,31,39] (6, 17%), and socioeconomic indicators, such as education level[30,37,39] (3, 9%), employment status [18,36,37] (3, 9%), or type of insurance[37] (1, 3%), as illustrated in Figure 4. Studies reporting on related predictive factors in the data were census region[‖] (24, 69%), race or ethnicity[#] (23, 66%), and health disparity population[18–20,25,29,31,38,49,51] (9, 26%). Racial and ethnic populations spanned American Indian or Alaskan Native, Asian, Black, Chinese, Hispanic, Native Hawaiian or Pacific Islander, non-Hispanic Black, non-Hispanic White, Mexican American, Spanish descent, White, and multiracial races or ethnicities. Other health disparity populations included were persons with disabilities. Overall, a significant proportion of the studies did not consider all the relevant patient- and community-level SDoH features that may influence CVD outcomes, nor did they incorporate diversity in model training data to ensure an appropriate representation of groups from health disparity or medically underserved populations. Table 2 organizes demographics and SDoH around clusters (age, sex, race or ethnicity, preexisting conditions, behavioral trends, exercise, and diet) and what insights they provide for CVD risks and outcomes.

Discussion

Our targeted review of the current literature shows that there are limited published studies leveraging advanced analytics and ML models for analyzing and predicting CVD risk and outcomes. Of the published studies, race and relevant SDoH factors known to influence CVD risk, progression, and outcomes are often missing in the analyses or model construction. While most (97%) of the reviewed papers included sex as a biologic variable,

[‡] *References 17-19, 21, 22, 24, 27, 28, 32, 33, 35-37, 39, 40, 47-49, 51.*

[§] *References 17, 18, 21, 23, 25, 26, 28-36, 38-43, 45-51.*

[‖] *References 18, 20, 23, 27-39, 41-43, 46, 48-51.*

[#] *References 18-20, 22, 25, 28-30, 32, 33, 36-40, 42, 43, 45, 49-51.*

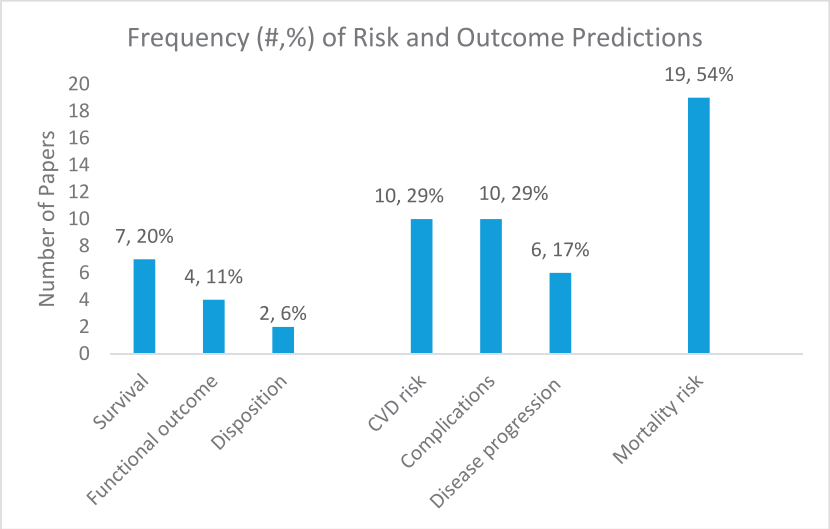

Figure 3. Risk and outcome predictions

only 66% of them included race or ethnicity variables, and 69% considered census region variables.

Potential biases of the use of race or ethnicity in AI and ML methodologies may occur for several reasons: (1) data that are used to train the models may be discriminatory or unrepresentative for people of color, women, or other marginalized groups, and (2) bias may be introduced, either intentionally or unintentionally, into the outcomes that the AI and ML models are predicting because models are built and used by humans. The potential harm of AI and ML methodologies that do not incorporate race or ethnicity is further compounded by the health information technology (IT) industry's lack of representation

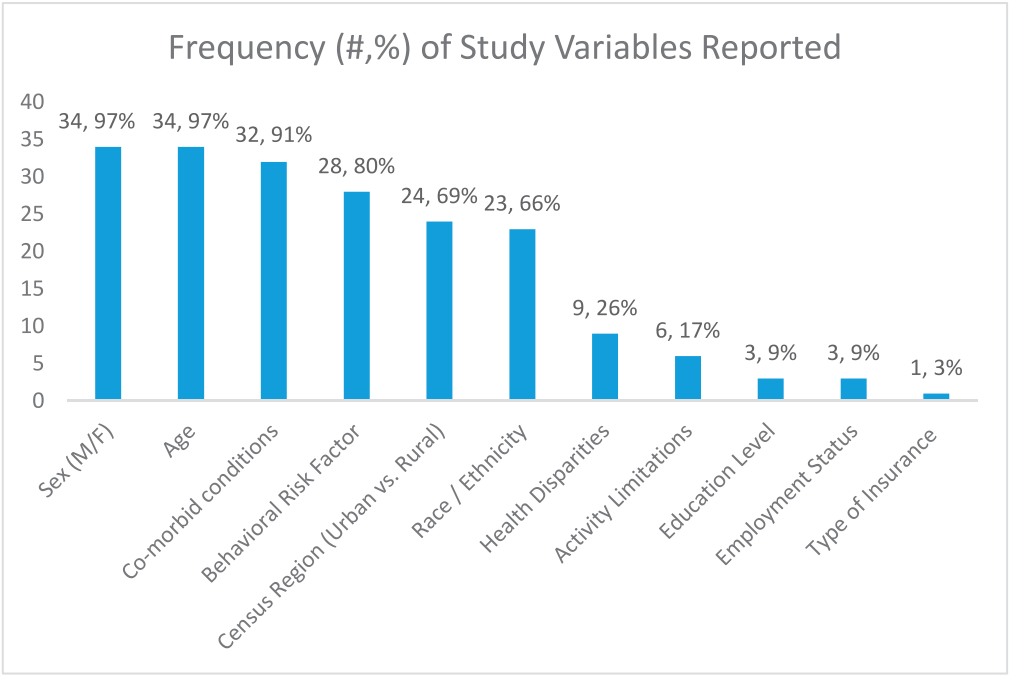

Figure 4. Frequency of study variables reported

Table 2. Key references exploring demographics and SDoH factors and their relationships for CVD risks and outcomes

Demographics and SDoH category related to disparities	References
Age	Age as a cardiovascular risk factor 2012[52]
Assigned sex at birth	The Framingham Study: historical insight on the effect of cardiovascular risk factors in men vs women 2002[53]
Race or ethnicity	Subclinical myocardial injury and cardiovascular mortality: racial differences in prevalence and risk (from the third National Health and Nutrition Examination survey 2021)[4]
	South Asian ethnicity as a risk factor for coronary heart disease 2020[54]
	Trends in racial or ethnic and nativity disparities in cardiovascular health among adults without prevalent cardiovascular disease in the United States, 1988–2014[55]
Preexisting conditions	Quality-adjusted life years (QALY) for 15 chronic conditions and combinations of conditions among US adults aged 65 and older 2018[56]
	Associations of multiple chronic health conditions with active life expectancy in the United States 2016[57]
	Multiple chronic conditions and life expectancy: a life table analysis 2014[58]
Health behaviors	Lifestyle-based prediction model for the prevention of CVD: the Healthy Heart Score 2014[59]
Income and employment	Disparities in absolute cardiovascular risk, metabolic syndrome, hypertension, and other risk factors by income within racial/ethnic groups among middle-aged and older US people 2021[60]
	Self-employment and cardiovascular risk in the US general population 2020[61]
	Income disparities in absolute cardiovascular risk and cardiovascular risk factors in the United States, 1999–2014[5]
Education	Modifiable risk factors, cardiovascular disease, and mortality in 155,722 individuals from 21 high-income, middle-income, and low-income countries (PURE): a prospective cohort study[62]
	Education, race or ethnicity, and causes of premature mortality among middle-aged adults in 4 US urban communities: results from CARDIA, 1985–2017[63]
	Contribution of major diseases to disparities in mortality 2002[64]
Dwelling location	Pre-pregnancy hypertension among women in rural and urban areas of the United States 2020[65]

of people who understand and can address these challenges.

Very few studies incorporated data on relevant social attributes, such as education level, unemployment, or health insurance status. Most studies used downstream measures (eg, health-related behaviors and receipt of recommended medical care) and lacked inclusion of upstream SDoH measures (eg, social disadvantage and racism). The notable lack of related SDoH variables in predictive models for CVD risk and outcomes is of particular concern. While it is well documented that a strong relationship exists between socioeconomic contexts and the health of individuals and populations, the use of SDoH variables in predictive models is important for examining possible causal pathways to inform care management for CVD. Limitations in the use of SDoH variables in predictive models may include the use of some self-reported variables, which may introduce a risk of recall bias, and the lack of availability of some factors in a data set.

Even though disparities in CVD prevention and treatment have been evident in the literature for decades,[66,67] existing predictive analytics or AI models often do not fully consider and explain the exacerbated risk facing disadvantaged populations. For example, the American College of Cardiology/American Heart Association Pooled Cohort Equations Risk Model systematically underpredicts atherosclerotic CVD among disadvantaged populations; however, adjusting the neighborhood disadvantage index can dramatically improve the risk prediction.[68]

Additional observations regarding this limited review include the following:

1. While there is increased awareness of the associations between patient-level social, behavioral, and economic factors as well as community-level factors of food access or neighborhood social capital on CVD risk, progression, and adverse outcomes, studies that predict risk, disease progression, morbidity survival, functionality, and survival outcomes do not include these features in their analyses.
2. Many studies relied heavily on the clinical data sets, research databases, or real-world evidence data sources (eg, blood pressure, triglyceride levels, artery wall thickness, comorbid conditions, adherence to medications, and symptomatology) that are obtained from either electronic health records (EHRs) or administrative claims to make these predictions.
3. Most studies that incorporate advanced analytics or ML also incorporate static social variables in addition to the clinical features, which may change over time and thus affect the prediction or risk.
4. An enhanced data architecture and expanded infrastructure, as well as a data analytics platform, are needed to

advance the technologies to achieve the potential of optimal CVD risk prediction.

5. Ethical frameworks for inclusion of SDoH variables into risk and predictive algorithms, as well as standardization of SDoH measures for feature engineering, may be needed to advance the accuracy of risk and predictive models for cardiovascular disease.

Hammond et al[69] examined the 2016–2017 Medicare Current Beneficiary Survey and found that the addition of a panel of SDoH measures to traditional risk adjustment models improved the prediction accuracy of clinical risk models for CVD hospitalization, annual cost, and death among racial and ethnic minorities. Models without SDoH significantly underpredicted all-cause and cardiovascular hospitalizations by 20% and 70%, respectively. Although further research is needed to explore whether SDoH plays the most crucial role in risk prediction for CVD or other chronic disease hospitalization and mortality, this study provided meaningful evidence of the importance of incorporating SDoH in predictive analytics, including AI. As more health systems use EHRs to manage chronic illness with the goal of improving health outcomes, AI and ML techniques have advanced simultaneously to provide smart solutions for clinical decision support and provider or patient education. Incorporating SDoH parameters in AI or ML algorithms is vital for establishing an equitable and unbiased system.

It is critical for advanced analytical technologies and AI models to incorporate equity, diversity, and inclusion throughout the data capture, model design, development, and implementation phases. To do this, first we need to incorporate standardized collection and input of race or ethnicity, socioeconomic status, and other SDoH measures in all systems that collect health data.[70] Because most AI and ML algorithms and models are built on existing health information infrastructure, including extensive use of EHR data, we need to identify deficiencies in the existing health IT system's ability to support health equity measures and adopt stakeholder-centered approaches to address these deficiencies.[71] Additionally, it was suggested that SDoH measures (eg, substance abuse and housing status) captured in patient records (both structured and unstructured, such as clinical notes) may exhibit changes over time due to data quality and other issues.[72] Multiple instances of SDoH documentation may be needed. Also, natural language processing tools could be used to extract SDoH information from clinical notes to further verify the health equity status and changes over time.

In addition to systemic complexity, providers often find it is difficult to obtain SDoH information from patients and retrieve SDoH information from EHRs, and some may be unaware of its consequences on low quality of care among populations affected by health disparities.[73] Medical education and awareness efforts to encourage incorporating SDoH into predictive analytics for disease management are hence warranted.

The advancement of data science and high-performance computing have initiated the AI era. Using advanced analytics and AI approaches for predictive modeling provides new opportunities to better understand CVD prevention and treatment at the individual and population levels. However, it also raises the health equity concern given the many lessons learned, including most recently that racial or ethnic minorities and other vulnerable populations suffered disproportional burden due to barriers in access and high risks of chronic comorbidities during the COVID-19 pandemic. Advanced analytics and AI approaches that do not incorporate or consider demographic and SDoH factors will likely lead to biased results, which may in turn lead to undesirable health outcomes. Thus, developing novel methods and intervention research using AI to address minority health and health disparities is an important component for enhancing the use of EHR to improve health care quality. As an example, the US National COVID Cohort Collaborative, with its additional linkages to many social, environmental, and behavioral public data sets, has provided a valuable open-access resource for the scientific community. Furthermore, the National Institutes of Health (NIH) established the AIM-AHEAD program to increase the participation and representation of underrepresented researchers and communities in AI or ML model development.[74] SDoH should be considered a major influencer in biomedical research.[75,76] Both precision medicine and precision public health need to incorporate demographic and SDoH information in clinical or policy decision making in order to identify the right intervention for the right person in the right community at the right dose and the right time.[77] For the first time in history, the NIH has recommended its COVID-19 funding opportunity announcements to incorporate the PhenX Social Determinants of Health Assessments Collection.[76,78] The collection includes not only individual SDoH (eg, access to health service, discrimination, English proficiency, job insecurity, and wealth) but also structural SDoH (eg, air quality, environmental justice, residential segregation, and social vulnerability).

Our review is limited to CVD literature and does not capture all the evidence from other chronic diseases and conditions using advanced analytics and AI approaches. Moreover, many international countries have pioneered using AI for CVD prevention and treatment. Their experience

and lessons learned could be useful for US and Canadian populations.

Conclusions

Health equity is the foundation for a strong and resilient society. However, currently, advanced analytics and AI approaches do not consistently factor in SDoH and health disparities in most cases. If advanced analytical technologies and AI models could effectively capture and incorporate these factors, resulting insights may improve how health equity is addressed in cardiovascular health. More innovative means by which to extract and standardize SDoH information from EHRs and other data sources are needed.

Disclaimer

The contents and views expressed in this report are those of the authors and do not necessarily reflect the official views of the National Institutes of Health, the Department of Health and Human Services, the US government, or the affiliated institutions.

Conflict of Interest

No conflicts of interest reported by authors.

Author Contributions

Research concept and design: Snowdon, Scheufele, Pritts, Zhang, Dankwa-Mullan; Acquisition of data: Scheufele, Pritts, Le, Zhang; Data analysis and interpretation: Snowdon, Scheufele, Pritts, Le, Zhang, Dankwa-Mullan; Manuscript draft: Snowdon, Scheufele, Pritts, Mensah, Zhang, Dankwa-Mullan; Statistical expertise: Le, Mensah, Zhang; Administrative: Snowdon, Scheufele, Pritts, Le, Mensah, Zhang, Dankwa-Mullan; Supervision: Snowdon, Dankwa-Mullan

References

1. Liem DA, Murali S, Sigdel D, et al. Phrase mining of textual data to analyze extracellular matrix protein patterns across cardiovascular disease. *Am J Physiol Heart Circ Physiol*. 2018;315(4):H910–H924. https://doi.org/10.1152/ajpheart.00175.2018
2. Arruda-Olson AM, Afzal N, Priya Mallipeddi V, et al. Leveraging the electronic health record to create an automated real-time prognostic tool for peripheral arterial disease. *J Am Heart Assoc.* 2018;7(23):e009680. https://doi.org/10.1161/JAHA.118.009680
3. Virani SS, Alonso A, Aparicio HJ, et al. Heart disease and stroke statistics—2021 update: a report from the American Heart Association. *Circulation*. 2021;143(8):e254–e743. https://doi.org/10.1161/CIR.0000000000000950
4. Broughton ST, Ahmad M, Soliman EZ, Magnani JW. Subclinical myocardial injury and cardiovascular mortality: racial differences in prevalence and risk (from the third National Health and Nutrition Examination Survey). *Ann Noninvasive Electrocardiol*. 2021;26(4):e12827. https://doi.org/10.1111/anec.12827
5. Odutayo A, Gill P, Shepherd S, et al. Income disparities in absolute cardiovascular risk and cardiovascular risk factors in the United States, 1999–2014. *JAMA Cardiol*. 2017;2(7):782–790. https://doi.org/10.1001/jamacardio.2017.1658
6. Hertz RP, Unger AN, Cornell JA, Saunders E. Racial disparities in hypertension prevalence, awareness, and management. *Arch Intern Med*. 2005;165(18):2098–2104. https://doi.org/10.1001/archinte.165.18.2098
7. Kent JA, Patel V, Varela NA. Gender disparities in health care. *Mt Sinai J Med*. 2012;79(5):555–559. https://doi.org/10.1002/msj.21336
8. Kanchi R, Perlman SE, Chernov C, et al. Gender and race disparities in cardiovascular disease risk factors among New York City Adults: New York City Health and Nutrition Examination Survey (NYC HANES) 2013–2014. *J Urban Health*. 2018;95(6):801–812. https://doi.org/10.1007/s11524-018-0287-x
9. DeCamp M, Lindvall C. Latent bias and the implementation of artificial intelligence in medicine. *J Am Med Inform Assoc*. 2020;27(12):2020–2023. https://doi.org/10.1093/jamia/ocaa094
10. Marcelin JR, Siraj DS, Victor R, Kotadia S, Maldonado YA. The impact of unconscious bias in healthcare: how to recognize and mitigate it. *J Infect Dis*. 2019;220(suppl 2):S62–S73. https://doi.org/10.1093/infdis/jiz214
11. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447–453. https://doi.org/10.1126/science.aax2342
12. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. *Ann Intern Med*. 2018;169(12):866–872. https://doi.org/10.7326/M18-1990
13. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet*. 2019;51(4):584–591. https://doi.org/10.1038/s41588-019-0379-x
14. Pearson TA, Califf RM, Roper R, et al. Precision health analytics with predictive analytics and implementation research: JACC state-of-the-art review. *J Am Coll Cardiol*. 2020;76(3):306–320. https://doi.org/10.1016/j.jacc.2020.05.043
15. Goff DC Jr, Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation*. 2014;129(25 suppl 2):S49–S73. https://doi.org/10.1161/01.cir.0000437741.48606.98
16. Matheny M, McPheeters ML, Glasser A, et al. Systematic review of cardiovascular disease risk assessment tools. Evidence Synthesis No. 85. Rockville, MD: Agency for Healthcare Research and Quality; 2011.
17. Acher C, Acher CW, Havlena J, Wynn M. Advances in treatment and long-term survival in patients with descending thoracic aortic aneurysms treated at a single tertiary center from 1984 to 2014. *Ann Vasc Surg*.2019;58:190–197. https://doi.org/10.1016/j.avsg.2018.10.043
18. Ahmad MI, Dutta A, Anees MA, Soliman EZ. Interrelations between serum uric acid, silent myocardial infarction, and mortality in the general population. *Am J Cardiol*. 2019;123(6):882–888. https://doi.org/10.1016/j.amjcard.2018.12.016
19. Alawieh A, Starke RM, Chatterjee AR, et al. Outcomes of endovascular thrombectomy in the elderly: a "real-world" multicenter study. *J Neurointerv Surg*. 2019;11(6):545–553. https://doi.org/10.1136/neurintsurg-2018-014289
20. Alawieh A, Vargas J, Fargen KM, et al. Impact of procedure time on outcomes of thrombectomy for stroke. *J Am Coll Cardiol*. 2019;73(8):879–890. https://doi.org/10.1016/j.jacc.2018.11.052
21. Allen KB, Chhatriwalla AK, Cohen D, et al. Transcarotid versus transapical and transaortic access for transcatheter aortic valve replacement. *Ann Thorac Surg*. 2019;108(3):715–722. https://doi.org/10.1016/j.athoracsur.2019.02.007
22. Anadani M, Alawieh A, Vargas J, Chatterjee AR, Turk A, Spiotta A. First attempt recanalization with ADAPT: rate, predictors, and outcome. *J Neurointerv Surg*. 2019;11(7):641–645. https://doi.org/10.1136/neurintsurg-2018-014294
23. Ban VS, El Ahmadieh TY, Aoun SG, et al. Prediction of outcomes for ruptured aneurysm surgery. *Stroke*. 2019;50(3):595–601. https://doi.org/10.1161/strokeaha.118.023771
24. Bhattacharya M, Lu DY, Kudchadkar SM, et al. Identifying ventricular arrhythmias and their predictors by applying machine learning methods to electronic health records in patients with hypertrophic cardiomyopathy (HCM-VAr-Risk Model). *Am J Cardiol*. 2019;123(10):1681–1689. https://doi.org/10.1016/j.amjcard.2019.02.022
25. Boitano LT, Iannuzzi JC, Tanious A, et al. Preoperative predictors of discharge destination after endovascular repair of abdominal aortic aneurysms. *Ann Vasc Surg*. 2019;57:109–117. https://doi.org/10.1016/j.avsg.2018.12.058
26. Chang A, Ricci B, Grory BM, et al. Cardiac biomarkers predict large vessel occlusion in patients with ischemic stroke. *J Stroke Cerebrovasc Dis*. 2019;28(6):1726–1731. https://doi.org/10.1016/j.jstrokecerebrovasdis.2019.02.013
27. Chatterjee S, LeMaire SA, Amarasekara HS, et al. Early-stage acute kidney injury adversely affects thoracoabdominal aortic aneurysm repair outcomes. *Ann Thorac Surg*. 2019;107(6):1720–

1726. https://doi.org/10.1016/j.athoracsur.2018.11.049
28. Dudum R, Dzaye O, Mirbolouk M, et al. Coronary artery calcium scoring in low risk patients with family history of coronary heart disease: validation of the SCCT guideline approach in the coronary artery calcium consortium. *J Cardiovasc Comput Tomogr*. 2019;13(3):21–25. https://doi.org/10.1016/j.jcct.2019.03.012
29. Gepner AD, McClelland RL, Korcarz CE, et al. Carotid artery displacement and cardiovascular disease risk in the Multi-Ethnic Study of Atherosclerosis. *Vasc Med*. 2019;24(5):405–413. https://doi.org/10.1177/1358863x19853362
30. Huntley GD, Tecson KM, Sodhi S, et al. Cardiac denial and expectations associated with depression in adults with congenital heart disease. *Am J Cardiol*. 2019;123(12):2002–2005. https://doi.org/10.1016/j.amjcard.2019.03.011
31. Kasner SE, Siegler JE, Zamzam A, Kleindorfer D. Expanding eligibility in stroke prevention trials to patients with early disability. *J Stroke Cerebrovasc Dis*. 2019;28(8):2268–2272. https://doi.org/10.1016/j.jstrokecerebrovasdis.2019.05.016
32. Kilic A, Acker MA, Gleason TG, et al. Clinical outcomes of mitral valve reoperations in the United States: an analysis of the Society of Thoracic Surgeons National Database. *Ann Thorac Surg*. 2019;107(3):754–759. https://doi.org/10.1016/j.athoracsur.2018.08.083
33. Kumar A, Patel DR, Wolski KE, et al. Baseline fasting plasma insulin levels predict risk for major adverse cardiovascular events among patients with diabetes and high-risk vascular disease: insights from the ACCELERATE trial. *Diab Vasc Dis Res*. 2019;16(2):171–177. https://doi.org/10.1177/1479164119827604
34. Lal BK, Roubin GS, Jones M, et al. Influence of multiple stents on periprocedural stroke after carotid artery stenting in the Carotid Revascularization Endarterectomy versus Stent Trial (CREST). *J Vasc Surg*. 2019;69(3):800–806. https://doi.org/10.1016/j.jvs.2018.06.221
35. Lee DS, Lee JS, Schull MJ, et al. Prospective validation of the emergency heart failure mortality risk grade for acute heart failure. *Circulation*. 2019;139(9):1146–1156. https://doi.org/10.1161/circulationaha.118.035509
36. Lee JH, Rizvi A, Hartaigh B, et al. The predictive value of coronary artery calcium scoring for major adverse cardiac events according to renal function (from the Coronary Computed Tomography Angiography Evaluation for Clinical Outcomes: An International Multicenter [CONFIRM] Registry). *Am J Cardiol*. 2019;123(9):1435–1442. https://doi.org/10.1016/j.amjcard.2019.01.055
37. Lekoubou A, Bishu KG, Ovbiagele B. Association of prevalent stroke with hospitalization for seizure: patterns and prognoses. *J Stroke Cerebrovasc Dis*. 2019;28(11):104344. https://doi.org/10.1016/j.jstrokecerebrovasdis.2019.104344
38. Li DL, Quispe R, Madan N, Zhang L, Taub CC. A risk score for predicting atrial fibrillation in individuals with preclinical diastolic dysfunction: a retrospective study in a single large urban center in the United States. *BMC Cardiovasc Disord*. 2019;19(1):47. https://doi.org/10.1186/s12872-019-1024-4
39. Madhavan M, Holmes DN, Piccini JP, et al. Association of frailty and cognitive impairment with benefits of oral anticoagulation in patients with atrial fibrillation. *Am Heart J*. 2019;211:77–89. https://doi.org/10.1016/j.ahj.2019.01.005
40. Nejim B, Zarkowsky D, Hicks CW, Locham S, Dakour Aridi H, Malas MB. Predictors of in-hospital adverse events after endovascular aortic aneurysm repair. *J Vasc Surg*. 2019;70(1):80–91. https://doi.org/10.1016/j.jvs.2018.10.093
41. Noori VJ, Healey CT, Eldrup-Jorgensen J, et al. Comparison of major adverse event rates after elective endovascular aneurysm repair in New England using a novel measure of complication severity. *J Vasc Surg*. 2019;70(1):74–79. https://doi.org/10.1016/j.jvs.2018.10.055
42. O'Donnell TFX, Wade JE, Liang P, et al. Endovascular aneurysm repair in patients over 75 is associated with excellent 5-year survival, which suggests benefit from expanded screening into this cohort. *J Vasc Surg*. 2019;69(3):728–737. https://doi.org/10.1016/j.jvs.2018.06.205
43. Polak JF, Herrington D, O'Leary DH. Associations of edge-detected and manual-traced common carotid artery intima-media thickness with incident peripheral artery disease: the Multi-Ethnic Study of Atherosclerosis. *Vasc Med*. 2019;24(4):306–312. https://doi.org/10.1177/1358863x19835925
44. Rocha EA, Ji R, Ay H, et al. Reduced ischemic lesion growth with heparin in acute ischemic stroke. *J Stroke Cerebrovasc Dis*. 2019;28(6):1500–1508. https://doi.org/10.1016/j.jstrokecerebrovasdis.2019.03.016
45. Sakr AE, Fraser GE, Doctorian TP, et al. Predictors of systolic heart failure and mortality following orthotopic liver transplantation: a single-center cohort. *Transplant Proc*. 2019;51(6):1950–1955. https://doi.org/10.1016/j.transproceed.2019.04.063
46. Samad MD, Ulloa A, Wehner GJ, et al. Predicting survival from large echocardiography and electronic health record datasets: optimization with machine learning. *JACC Cardiovasc Imaging*. 2019;12(4):681–689. https://doi.org/10.1016/j.jcmg.2018.04.026
47. Shah M, Patil S, Patnaik S, et al. Outcomes in cardiogenic shock from acute coronary syndrome depending on severity of obesity. *Am J Cardiol*. 2019;123(8):1267–1272. https://doi.org/10.1016/j.amjcard.2019.01.010
48. Shannon AH, Mehaffey JH, Cullen JM, et al. Preoperative beta blockade is associated with increased rates of 30-day major adverse cardiac events in critical limb ischemia patients undergoing infrainguinal revascularization. *J Vasc Surg*. 2019;69(4):1167–1172.e1. https://doi.org/10.1016/j.jvs.2018.07.077
49. Sharma A, Sun JL, Lokhnygina Y, et al. Patient phenotypes, cardiovascular risk, and ezetimibe treatment in patients after acute coronary syndromes (from IMPROVE-IT). *Am J Cardiol*. 15 2019;123(8):1193–1201. https://doi.org/10.1016/j.amjcard.2019.01.034
50. Toth PP, Fazio S, Wong ND, Hull M, Nichols GA. Risk of cardiovascular events in patients with hypertriglyceridaemia: a review of real-world evidence. *Diabetes Obes Metab*. 2020;22(3):279–289. https://doi.org/10.1111/dom.13921
51. Wang H, Li Y, Ning H, Wilkins J, Lloyd-Jones D, Luo Y. Using machine learning to integrate socio-behavioral factors in predicting cardiovascular-related mortality risk. *Stud Health Technol Inform*. 2019;264:433–437. https://doi.org/10.3233/shti190258
52. Dhingra R, Vasan RS. Age as a risk factor. *Med Clin North Am*. 2012;96(1):87–91. https://doi.org/10.1016/j.mcna.2011.11.003.
53. Kannel WB. The Framingham Study: historical insight on the impact of cardiovascular risk factors in men versus women. *J Gend Specif Med*. 2002;5(2):27–37.
54. Pursnani S, Merchant M. South Asian ethnicity as a risk factor for coronary heart disease. *Atherosclerosis*. 2020 Dec;315:126–130. https://doi.org/10.1016/j.atherosclerosis.2020.10.007
55. Brown AF, Liang LI, Vassar SD, et al. Trends in racial/ethnic and nativity disparities in cardiovascular health among adults without prevalent cardiovascular disease in the United States, 1988 to 2014. *Ann Intern Med*. April 17, 2018. https://doi.org/10.7326/M17-0996
56. Jia H, Lubetkin EI, Barile JP, et al. Quality-adjusted life years (QALY) for 15 chronic conditions and combinations of conditions among US adults aged 65 and older. *Med Care*. 2018;56(8):740–746.
57. Laditka JN, Laditka SB. Associations of multiple chronic health conditions with active life expectancy in the United States. *Disabil Rehabil*. 2016;38(4):354–361.
58. DuGoff EH, Canudas-Romo V, Buttorff C, Leff B, Anderson GF. Multiple chronic conditions and life expectancy: a life table analysis. *Med Care*. 2014;52(8):688–694.
59. Chiuve SE, Cook NR, Shay CM et al. Lifestyle-based prediction model for the prevention of CVD: the Healthy Heart Score 2014. *J Am Heart Assoc.* 2014;3(6)e000954. https://doi.org/10.1161/JAHA.114.000954.
60. Al Kibria GM, Crispen R, Chowdhury MAB, Rao N, Stennett C. Disparities in absolute cardiovascular risk, metabolic syndrome, hypertension, and other risk factors by income within racial/ethnic groups among middle-aged and older US people 2021. [published online ahead of print March 5, 2021]. *J Hum Hypertens.* https://doi.org/10.1038/s41371-021-00513-8 Erratum 2021;35(7):645. https://doi.org/10.1038/s41371-021-00527-2
61. Krittanawong C, Kumar A, Wang Z, Baber U, Bhatt DL. Self-employment and cardiovascular risk in the US general population 2020. *Int J Cardiol Hypertens.* 2020;6:100035. https://doi.org/10.1016/j.ijchy.2020.100035
62. Yusuf S, Joseph P, Rangarajan S, et al. Modifiable risk factors, cardiovascular disease, and mortality in 155 722 individuals from 21 high-income, middle-income, and low-income countries (PURE): a prospective cohort study.*Lancet*.

2020;395(10226):795–808. https://doi.org/10.1016/S0140-6736(19)32008-2 Erratum 395 (10226):784. https://doi.org/10.1016/S0140-6736(19)32282-2
63. Roy B, Kiefe CI, Jacobs DR, et al. Education, race/ethnicity, and causes of premature mortality among middle-aged adults in 4 US urban communities: results from CARDIA, 1985-2017.*Am J Public Health*. 2020;110(4):530–536. https://doi.org/10.2105/AJPH.2019.305506
64. Wong MD, Shapiro MF, Boscardin WJ, Ettner SL. Contribution of major diseases to disparities in mortality. *N Engl J Med.* 2002;347(20):1585–1592.
65. Cameron NA, Molsberry R, Pierce JB, et al. Pre-pregnancy hypertension among women in rural and urban areas of the United States 2020. *J Am Coll Cardiol*. 2020;76(22):2611–2619. https://doi.org/10.1016/j.jacc.2020.09.601
66. Mensah GA, Mokdad AH, Ford ES, Greenlund KJ, Croft JB. State of disparities in cardiovascular health in the United States. *Circulation*. 2005;111(10):1233–1241. https://doi.org/10.1161/01.CIR.0000158136.76824.04
67. Shahu A, Herrin J, Dhruva SS, et al. Disparities in socioeconomic context and association with blood pressure control and cardiovascular outcomes in ALLHAT. *J Am Heart Assoc.* 2019;8(15):e012277. https://doi.org/10.1161/JAHA.119.012277
68. Dalton JE, Perzynski AT, Zidar DA, et al. Accuracy of cardiovascular risk prediction varies by neighborhood socioeconomic position: a retrospective cohort study. *Ann Intern Med*. 2017;167(7):456–464. https://doi.org/10.7326/M16-2543
69. Hammond G, Johnston K, Huang K, Joynt Maddox KE. Social determinants of health improve predictive accuracy of clinical risk models for cardiovascular hospitalization, annual cost, and death. *Circ Cardiovasc Qual Outcomes*. 2020;13(6):e006752. https://doi.org/10.1161/CIRCOUTCOMES.120.006752
70. Zhang XZ, Perez-Stable EJ, Bourne PE, et al. Big data science: opportunities and challenges to address minority health and health disparities in the 21st century. *Ethnic Dis*. 2017;27(2):95–106. https://doi.org/10.18865/ed.27.2.95
71. Zhang X, Hailu B, Tabor DC, et al. Role of health information technology in addressing health disparities: patient, clinician, and system perspectives. *Med Care*. 2019;57(suppl 6 2: S115–S120. https://doi.org/10.1097/MLR.0000000000001092
72. Feller DJ, Zucker J, Walk OBD, Yin MT, Gordon P, Elhadad N. Longitudinal analysis of social and behavioral determinants of health in the EHR: exploring the impact of patient trajectories and documentation practices. *AMIA Annu Symp Proc*. 2019;2019:399–407.
73. Gottlieb LM, Tirozzi KJ, Manchanda R, Burns AR, Sandel MT. Moving electronic medical records upstream: incorporating social determinants of health. *Am J Prev Med*. 2015;48(2):215–218. https://doi.org/10.1016/j.amepre.2014.07.009
74. National Institutes of Health. *Artificial Intelligence/Machine Learning Consortium to Advance Health Equity and Researcher Diversity (AIM-AHEAD)*. https://datascience.nih.gov/artificial-intelligence/aim-ahead
75. Collins FS. Creative minds: considering the social determinants of health. *NIH Director's Blog*. December 28, 2016. https://directorsblog.nih.gov/2016/05/12/creative-minds-considering-the-social-determinants-of-health/
76. Figueroa JF, Frakt AB, Jha AK. Addressing social determinants of health: time for a polysocial risk Score. *JAMA*. 2020;323(16):1553–1554. https://doi.org/10.1001/jama.2020.2436
77. Pearson TA, Califf RM, Roper R, et al. Precision health analytics with predictive analytics and implementation research: JACC state-of-the-art review. *J Am Coll Cardiol*. 2020;76(3):306–320. https://doi.org/10.1016/j.jacc.2020.05.043
78. National Institutes of Health. *Notice Announcing Availability of Data Harmonization Tools for Social Determinants of Health (SDOH) via the PhenX Toolkit*. https://grants.nih.gov/grants/guide/notice-files/NOT-MD-21-003.html, https://www.phenxtoolkit.org/collections/view/6