

UCSF

UC San Francisco Previously Published Works

Title

Heterogeneous Exposure Associations in Observational Cohort Studies: The Example of Blood Pressure in Older Adults

Permalink

<https://escholarship.org/uc/item/2q1259p2>

Journal

American Journal of Epidemiology, 189(1)

ISSN

0002-9262

Authors

Odden, Michelle C
Rawlings, Andreea M
Khodadadi, Abtin
[et al.](#)

Publication Date

2020-01-31

DOI

10.1093/aje/kwz218

Peer reviewed

Practice of Epidemiology

Heterogeneous Exposure Associations in Observational Cohort Studies: The Example of Blood Pressure in Older Adults

Michelle C. Odden*, Andreea M. Rawlings, Abtin Khodadadi, Xiaoli Fern, Michael G. Shlipak, Kirsten Bibbins-Domingo, Kenneth Covinsky, Alka M. Kanaya, Anne Lee, Mary N. Haan, Anne B. Newman, Bruce M. Psaty, and Carmen A. Peralta

* Correspondence to Dr. Michelle C. Odden, Department of Epidemiology and Population Health, School of Medicine, Stanford University, 259 Campus Drive, Stanford, CA 94305 (e-mail: modden@stanford.edu).

Initially submitted October 17, 2018; accepted for publication October 3, 2019.

Heterogeneous exposure associations (HEAs) can be defined as differences in the association of an exposure with an outcome among subgroups that differ by a set of characteristics. In this article, we intend to foster discussion of HEAs in the epidemiologic literature and present a variant of the random forest algorithm that can be used to identify HEAs. We demonstrate the use of this algorithm in the setting of the association between systolic blood pressure and death in older adults. The training set included pooled data from the baseline examination of the Cardiovascular Health Study (1989–1993), the Health, Aging, and Body Composition Study (1997–1998), and the Sacramento Area Latino Study on Aging (1998–1999). The test set included data from the National Health and Nutrition Examination Survey (1999–2002). The hazard ratios ranged from 1.25 (95% confidence interval: 1.13, 1.37) per 10-mm Hg increase in systolic blood pressure among men aged ≤ 67 years with diastolic blood pressure greater than 80 mm Hg to 1.00 (95% confidence interval: 0.96, 1.03) among women with creatinine concentration ≤ 0.7 mg/dL and a history of hypertension. HEAs have the potential to improve our understanding of disease mechanisms in diverse populations and guide the design of randomized controlled trials to control exposures in heterogeneous populations.

blood pressure; effect modification; epidemiologic methods; random forests

Abbreviations: BMI, body mass index; CHS, Cardiovascular Health Study; CI, confidence interval; DBP, diastolic blood pressure; HEA, heterogeneous exposure association; Health ABC, Health, Aging, and Body Composition; HR, hazard ratio; LDL, low-density lipoprotein; NHANES, National Health and Nutrition Examination Survey; SALSA, Sacramento Area Latino Study on Aging; SBP, systolic blood pressure.

There has recently been interest in the identification of heterogeneous treatment effects in clinical trials, defined as clinically important differences in the benefit or harm of a treatment among patients, on the basis of one or several patient characteristics (1). This interest is driven, in part, by precision medicine, which proposes “prevention and treatment strategies that take individual variability into account” (2, p. 793). Although the principal focus of precision medicine has been on genetic effect modifiers of treatment, nongenetic factors are also important determinants of the potential benefits and harms of a therapy. An essential component of medicine is knowledge about which patients may derive the most benefit from a given therapy and which

may derive no benefit, or even be harmed. Traditionally, this has been examined in either prespecified or post-hoc subgroup analyses in clinical trials. This can lead to multiple subgroup analyses which may overlap in patient populations and are underpowered and subject to spurious findings due to multiple comparisons (1). Of key importance, subpopulations are frequently defined by a single variable, such as age or sex, whereas heterogeneous treatment effects may be best identified using a multitude of characteristics in combination.

The discussion of heterogeneous treatment effects has been restricted to clinical trials; thus, we aim to expand this discussion to observational studies, where the goal is to iden-

tify meaningful differences in the effect of a hypothesized causal exposure on an outcome among study participants that differ by a set of characteristics (1, 3–5). This can be described as heterogeneous exposure associations (HEAs) and can be thought of as the observational analog of heterogeneous treatment effects. Effect modification is usually defined as a situation where “the effect of one variable on another differs across strata of a third” (6, p. 864). A key distinction is that HEAs may define effect modification based on multiple characteristics in combination. Observational epidemiology is frequently used to develop and test hypotheses regarding the effect of exposures in a population; thus, identification of HEAs could lead to more targeted intervention strategies and the evolution of precision population health. Additionally, the identification of HEAs could help explain why results differ across study populations and could be combined with reweighting methods to test the transportability of results across target populations (7).

Methods for evaluating effect modification by a set of variables can be cumbersome and limited by investigator-driven hypotheses. We propose a data-driven approach for the identification of HEAs in observational studies based on a variant of a random forest classification algorithm. The purpose of this algorithm is to identify multivariate (combinations of variables) modifiers of the effect of a hypothesized causal exposure on an outcome of interest. To simplify the presentation of this method, we do not address confounding of the relationship of interest, although this would be required for the identification of causal relationships (i.e., the exchangeability assumption), along with the positivity and consistency assumptions (8–13). When evaluated on a multiplicative scale, HEAs can be evaluated by testing whether the ratio measures of effect are equal:

$$\frac{E\left[Y|A = 1, \vec{Z} = 0\right]}{E\left[Y|A = 0, \vec{Z} = 0\right]} \neq \frac{E\left[Y|A = 1, \vec{Z} = 1\right]}{E\left[Y|A = 0, \vec{Z} = 1\right]}.$$

Here Y is the outcome of interest, A is the exposure of interest, and \vec{Z} is a multivariate effect modifier. We demonstrate the use of this algorithm in the setting of the association between blood pressure level and risk of death among older participants in several well-known observational studies. Optimal blood pressure control in older adults has been a recent source of controversy, and research has demonstrated that measures of functional status may distinguish populations who may benefit from lower blood pressure and those in whom lower blood pressure is associated with increased risk of death or adverse outcomes (14–17). We aimed to examine whether combinations of variables could identify subgroups of older adults in whom the association between systolic blood pressure (SBP) and all-cause mortality differed consistently across multiple cohort studies, and whether these patterns were observed in a separate test sample. In this article, we intend to enhance discussion of HEAs in the epidemiologic literature and to present a modification of a well-described machine-learning approach

that can be used to identify HEAs in observational studies. We also discuss some considerations for the design of studies aimed at identification of HEAs.

METHODS

Study population

To identify an HEA, we used both training and test data sets. In machine learning, it is common to develop the model in one data set (the training data) and then identify a separate data set with which to test the performance of the model derived in the training data. Sometimes this is done by partitioning the data into 2 or more data sets. Incorporation of this step helps prevent overfitting of the data; evaluation in a separate data set can help strengthen the evidence of the generalizability of the findings (18). In our study, we selected training and test data sets from completely different studies. The training set included pooled data from the baseline examination of 3 National Institutes of Health-funded cohort studies of older adults: the Cardiovascular Health Study (CHS) (19); the Health, Aging, and Body Composition (Health ABC) Study (20); and the Sacramento Area Latino Study on Aging (SALSA) (21).

The CHS evaluated risk factors for cardiovascular disease in the elderly (19). CHS participants were recruited from Forsyth County, North Carolina; Sacramento County, California; Washington County, Maryland; and Pittsburgh, Pennsylvania, in 1989–1990 for the original cohort and in 1992–1993 for a supplementary cohort designed to increase the number of black participants. Eligibility criteria included the following: 1) being aged ≥ 65 years; 2) not being institutionalized; 3) expecting to remain in one’s current community for 3 years or longer; 4) not being under active treatment for cancer; and 5) being able to give informed consent without requiring a proxy respondent.

The Health ABC Study was designed to examine the relationship between age-related changes in health and body composition and incident functional limitations in well-functioning black and white adults aged 70–79 years (20). Participants were recruited at 2 study sites—Pittsburgh, Pennsylvania, and Memphis, Tennessee—from a list of Medicare beneficiaries between April 1997 and June 1998. Inclusion criteria were 1) reported ability to walk one-quarter mile (0.4 km), climb 10 steps, and perform basic Activities of Daily Living without difficulty; 2) absence of life-threatening illness; and 3) planning to remain in the immediate geographical area for at least 3 years.

SALSA was a longitudinal cohort study of community-dwelling elderly Mexican Americans aged 60–101 years at baseline in 1998–1999 who resided in 6 counties in California’s Sacramento Valley and was designed to study risk factors for dementia (21). Inclusion criteria were being aged 60 years or older in 1998, being self-designated as Latino, and being a resident of the Sacramento Metropolitan Statistical Area and surrounding suburban and rural counties.

The test set included data from the National Health and Nutrition Examination Survey (NHANES), a nationally representative survey of the civilian, noninstitutionalized US population conducted by the National Center for Health

Statistics of the Centers for Disease Control and Prevention. This study included data from participants aged 60 years or older from 2 waves of the survey (1999–2000 and 2001–2002); 3,234 persons completed both the interview and the medical examination (22). We selected these waves because they are the only waves at which gait speed was measured in NHANES.

All-cause mortality

The outcome of interest for this analysis was all-cause mortality. Briefly, for the CHS and the Health ABC Study, all events were adjudicated by an outcome-assessment committee (23). Deaths were identified from household contacts and by review of obituaries, medical records, and death certificates, by periodic review of the National Death Index, and by review of the Centers for Medicare and Medicaid Services health-care-utilization database for hospitalizations; 100% ascertainment of mortality status was achieved. In SALSA, mortality was ascertained using online obituary surveillance, review of the Social Security Administration's Death Master File and the National Death Index, review of identifiable vital statistics data files from California, and interviews with family members (24). The National Center for Health Statistics has linked mortality data from NHANES to death certificate data in the National Death Index (22). Mortality data were available from the date of survey participation through December 31, 2006, based on a probabilistic match between NHANES and National Death Index certificate records.

Exposure

The exposure of interest was SBP, which was measured in all studies as the average of 2 or more measurements taken in the seated position with a sphygmomanometer.

Effect modifiers

We chose candidate subgroup variables as those that would potentially be assessed in a clinical setting. Variables included age, sex, race/ethnicity (Latino, black, white), height, weight, body mass index (BMI; calculated as weight (kg)/height (m)²), diastolic blood pressure (DBP), depressive symptoms (measured using the Center for Epidemiologic Studies Depression Scale), and levels of low-density lipoprotein (LDL) cholesterol, high-density lipoprotein cholesterol, total cholesterol, triglycerides, creatinine, fasting glucose, fasting insulin, and C-reactive protein. We also included history of hypertension, diabetes mellitus, coronary heart disease, cerebrovascular disease, and heart failure; use of antihypertensive medications, diabetes medications, statins, and nonsteroidal antiinflammatory drugs; self-rated health (excellent, very good, good, fair, or poor); any limitation in 1 or more Activities of Daily Living; and any limitation in 1 or more Instrumental Activities of Daily Living. Physical activity was assessed as a z score in all cohorts: kilocalories per week spent in physical activity (excluding chores) in the CHS, kilocalories per week spent in exercise and walking in the Health ABC Study, and hours per week

of leisure-time physical activity in SALSA. Gait speed was measured in the CHS and Health ABC Study and self-reported in SALSA and was categorized as slow (<0.8 m/second or easy/casual pace), medium (0.8–1.1 m/second or moderate pace), or fast (≥ 1.2 m/second or brisk/very brisk pace) walking or lack of completion of the gait test (self-report of never walking outdoors or being unable to walk). Cognitive function was assessed by means of the Modified Mini-Mental State Examination in CHS, Health ABC, and SALSA.

Analytical approach

We determined the characteristics of the training cohorts. Next, we applied a variant of the random forest algorithm developed specifically to identify HEAs. A random forest is a classic ensemble learning method for classification or regression that operates by constructing multiple decision trees using bootstrapped samples of the training set and aggregating the prediction results from the individual trees (25). In our variant, each node of the tree represents a variable and its corresponding cutpoint. The algorithm selects the candidate variable and cutpoint pair that maximizes the absolute difference between the Cox coefficients across the 2 subgroups:

$$\arg \max f(x),$$

where x is a variable defining subgroups $x > X$ and $x \leq X$, and

$$f(x) = \text{abs}((\hat{\beta}|x > X) - (\hat{\beta}|x \leq X)).$$

Additionally, β is the coefficient of SBP from a Cox proportional hazards model:

$$\lambda(t|\text{SBP}) = \lambda_0(t) \exp(\beta \times \text{SBP}).$$

Toward this goal, we incrementally build a decision tree by greedily selecting one variable and cutpoint at a time. At each step, we examine all variable and cutpoint pairs, each pair leading to a partition of the current population into 2 subgroups. For example, Figure 1 represents an example of a single decision tree that is constructed on a random 90% sample of men. Diastolic blood pressure data are used to define 2 subgroups of participants in whom the coefficient on SBP of one group (DBP ≤ 83 mm Hg) maximally differs from the coefficient in the other group (DBP > 83 mm Hg). Within the first group, DBP ≤ 83 mm Hg, a BMI cutpoint of 29.2 defines a group of participants (BMI ≤ 29.2) in whom the coefficient on SBP maximally differs from the coefficient among those with BMI > 29.2 , and so on.

To build the forest, we built each tree with 90% of the training set sampled without replacement. We built 100 trees each for the total sample, in women and men separately. Sex was the primary node in nearly all of the total sample runs, so primary results are presented sex-stratified. Given a tree, each path defines a subpopulation and is considered as an individual candidate pattern. Patterns that occurred in 20% or more of trees were externally evaluated in the NHANES test set in order to minimize the risk of overfitting. Machine-learning methods generally incorporate either

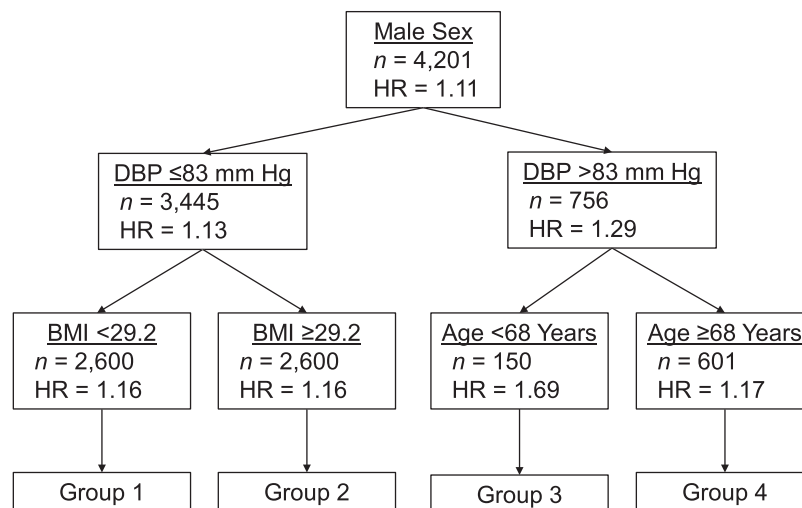


Figure 1. Example of a single decision tree constructed on a random 90% sample of men from 3 pooled cohorts (the Cardiovascular Health Study (1989–1993), the Health, Aging, and Body Composition Study (1997–1998), and the Sacramento Area Latino Study on Aging (1998–1999)), based on the hazard ratio (HR) for systolic blood pressure in a proportional hazards model of mortality. Body mass index (BMI) was calculated as weight (kg)/height (m)². DBP, diastolic blood pressure.

cross-validation or out-of-sample evaluation on a separate test data set to evaluate the models. This is in contrast to traditional statistical inference, which is based on hypothesis testing and typically uses *P* values or a related measure for inference. With machine learning, the number of comparisons does not impact the variables identified as being important—in this case, those identifying the heterogeneous subgroups (26). Our out-of-sample evaluation criteria required that the subgroup estimate differ from the overall estimate in the same direction for both the pooled cohorts (training data) and NHANES (test data)—for example, a weaker estimate in the subgroup than in the overall sample in both populations. We also assessed the magnitude of the difference between the coefficient for SBP in the subgroup and the coefficient in the overall population of women or men and noted those that varied by 25% or more.

Finally, our method allows for the estimate of the effect of SBP to vary by each branch (subgroup). Since effect modification can depend on the scale on which it is assessed (multiplicative or additive), we also calculated 10-year risk differences associated with a 10-mm Hg increase in SBP in the validated subpopulations using logistic regression models fitted in each subgroup (5, 27, 28).

The random forest algorithm was written using Python 3.4.5 (Python Software Foundation, Beaverton, Oregon), and all statistical analyses were run in Stata 14 (StataCorp LLC, College Station, Texas).

RESULTS

The descriptive characteristics of the training study population are listed in Table 1. Participants in the 3 cohort studies were, on average, in their 70s and mostly women. Participants in the CHS and Health ABC Study were predominantly white and black, and SALSA was comprised

entirely of Latino participants. Participants in the Health ABC Study were free of limitations in Activities of Daily Living or Instrumental Activities of Daily Living at baseline, by design. Participants in CHS had the highest 10-year mortality rate, followed by the Health ABC Study and SALSA, in descending order.

In the training set, the hazard ratio for the overall association between a 10-mm Hg increase in SBP and mortality was 1.08 (95% confidence interval (CI): 1.07, 1.10). Nearly all covariate combinations that occurred in at least 20% of trees included sex as the primary node. Among women in the training set, the hazard ratio for the association between a 10-mm Hg increase in SBP and mortality was 1.11 (95% CI: 1.10, 1.13) (Table 2). We observed 8 covariate and cutpoint combinations that occurred in at least 20% of trees in women in the training set. Of these, 6 met our out-of-sample verification criteria, and 4 differed by the overall coefficient in women by 25%. The 2 groups that met both of these criteria included women with an LDL cholesterol concentration less than or equal to 130 mg/dL and differed only by Modified Mini-Mental State Examination score of >80 points or ≤80 points; for a 10-mm Hg increase in SBP in the training set, hazard ratios were 1.15 (95% CI: 1.12, 1.18) and 1.04 (95% CI: 0.98, 1.11), respectively. This implies that among women with a lower LDL cholesterol level, SBP was associated with higher mortality risk in those with preserved cognitive function but not in those with impaired cognitive function.

Among men in the training set, the hazard ratio for the association between a 10-mm Hg increase in SBP and mortality was 1.05 (95% CI: 1.03, 1.07) (Table 3). We observed 6 covariate and cutpoint combinations that occurred in at least 20% of trees in the training set, and of these, 5 met our external validation criteria; 3 of these also met our criterion of varying from the overall coefficient by 25%. The subgroup-specific estimates for blood pressure in the

Table 1. Characteristics of Participants From 3 Observational Cohort Studies (the Cardiovascular Health Study (1989–1993), the Health, Aging, and Body Composition Study (1997–1998), and the Sacramento Area Latino Study on Aging (1998–1999))

Characteristic	CHS (n = 5,888)			Health ABC Study (n = 3,075)			SALSA (n = 1,789)					
	No. ^a	%	Mean (SD)	Median (IQR)	No.	%	Mean (SD)	Median (IQR)	No.	%	Mean (SD)	Median (IQR)
Age, years	3,393	57.6	72.8 (5.6)		1,584	51.5	73.6 (2.9)		1,038	58.3	70.6 (7.1)	
Female sex	4,925	83.6			1,794	58.3			0	0.0		
Race/ethnicity	924	15.7			1,281	41.7			0	0.0		
White	0	0.0			0	0.0			1,789	100.0		
Black	39	0.7			0	0.0			0	0.0		
Latino												
Other												
Education												
Less than high school	1,732	29.5			775	25.3			1,260	70.8		
High school graduate	1,620	27.6			1,000	32.6			225	12.6		
More than high school	2,519	42.9			1,292	42.1			294	16.5		
Self-rated health												
Excellent	790	13.4			419	13.6			113	6.6		
Very good	1,415	24.1			931	30.3			220	12.8		
Good	2,175	37.0			1,227	40.0			540	31.5		
Fair	1,256	21.4			471	15.3			669	39.1		
Poor	239	4.1			23	0.7			171	10.0		
Gait speed ^b												
Missing data	87	1.5			406	13.2			119	6.7		
Slow	2,914	49.5			41	1.3			494	27.6		
Medium	2,817	47.8			754	24.5			884	49.4		
Fast	70	1.2			1,874	60.9			292	16.3		
≥1 ADL limitation	476	9.0			0	0.0			226	13.0		
≥1 IADL limitation	1,513	28.1			0	0.0			1,307	75.3		

Table continues

Table 1. Continued

Characteristic	CHS (n = 5,886)			Health ABC Study (n = 3,075)			SALSA (n = 1,789)					
	No. ^a	%	Mean (SD)	Median (IQR)	No.	%	Mean (SD)	Median (IQR)	No.	%	Mean (SD)	Median (IQR)
Smoking status												
Never smoker	2,738	46.5			1,348	43.9			818	46.1		
Current smoker	2,444	41.6			1,404	45.7			754	42.5		
Former smoker	700	11.9			318	10.4			203	11.4		
CES-D score				4.0 (2.0–8.0)				3.0 (1.0–7.0)				6.0 (2.0–16.0)
Weight, pounds ^c			160.3 (32.4)				166.8 (33.1)					165.6 (33.0)
Height, inches ^d			64.9 (3.7)				65.4 (3.7)					62.7 (4.3)
Body mass index ^e			26.7 (4.7)				27.4 (4.8)					29.7 (5.9)
Laboratory measures												
Triglycerides, mg/dL				120.0 (92.0–164.0)				118.0 (88.0–164.0)				156.0 (111.0–227.0)
LDL cholesterol, mg/dL			130.0 (36.0)				122.0 (35.0)					123.0 (35.0)
HDL cholesterol, mg/dL			54.0 (16.0)				54.0 (17.0)					52.0 (14.0)
Total cholesterol, mg/dL			211.0 (39.0)				203.0 (39.0)					212.0 (40.0)
Fasting glucose, mg/dL				101.0 (94.0–112.0)				94.0 (87.0–106.0)				99.0 (88.0–123.0)
Insulin, mg/dL				13.0 (9.0–18.0)				6.9 (4.9–10.3)				8.8 (5.6–13.9)
C-reactive protein, mg/dL				2.5 (1.3–4.5)				1.7 (1.0–3.1)				3.3 (1.3–7.1)
Cystatin C, mg/L				1.0 (0.9–1.1)				1.0 (0.9–1.1)				1.0 (0.9–1.2)
Systolic BP, mm Hg			137.0 (22.0)				134.0 (21.0)					138.0 (19.0)
Diastolic BP, mm Hg			71.0 (12.0)				71.0 (12.0)					76.0 (11.0)
Physical activity z score ^f			0.0 (1.0)				0.0 (1.0)					0.0 (1.0)
3MS score				91.0 (87.0–94.0)				92.0 (86.0–96.0)				88.0 (80.0–93.0)
Chronic health conditions												
Hypertension	2,623	44.6			1,563	51.2			578	32.4		
Coronary heart disease	1,154	19.6			625	20.7			154	8.7		
Heart failure	275	4.7			95	3.1			52	2.9		
Diabetes mellitus	1,763	30.2			468	15.2			508	28.6		
Cerebrovascular disease	249	4.2			247	8.1			168	9.4		

Table continues

Table 1. Continued

Characteristic	CHS (n = 5,888)			Health ABC Study (n = 3,075)			SALSA (n = 1,789)					
	No. ^a	%	Mean (SD)	Median (IQR)	No.	%	Mean (SD)	Median (IQR)	No.	%	Mean (SD)	Median (IQR)
Medication use												
Diabetes medication	373	6.3			385	12.6			374	20.9		
Antihypertensive agents	2,789	47.4			1,672	54.6			761	42.6		
Statins	132	2.2			395	12.9			146	8.2		
Mortality over 10 years	2,070	35.2			988	32.1			500	27.9		

Abbreviations: ADL, Activities of Daily Living; BP, blood pressure; CES-D, Center for Epidemiologic Studies Depression Scale; CHS, Cardiovascular Health Study; HDL, high-density lipoprotein; Health ABC, Health, Aging, and Body Composition; IADL, Instrumental Activities of Daily Living; IQR, interquartile range; LDL, low-density lipoprotein; 3MS, Modified Mini-Mental State Examination; SALSA, Sacramento Area Latino Study on Aging; SD, standard deviation.

^a Number of participants.

^b Gait speed was measured in the CHS and Health ABC and self-reported in SALSA and was categorized as slow (<0.8 m/second or easy/casual pace), medium (0.8–1.1 m/second or moderate pace), or fast (≥ 1.2 m/second or brisk/very brisk pace) walking or lack of completion of the gait test (self-report of never walking outdoors or being unable to walk).

^c 1 pound = 0.45 kg.

^d 1 inch = 2.54 cm.

^e Weight (kg)/height (m)².

^f Physical activity was assessed as a z score in all cohorts: kilocalories per week spent in physical activity (excluding chores) in the CHS, kilocalories per week spent in exercise and walking in the Health ABC Study, and hours per week of leisure-time physical activity in SALSA.

Table 2. Association Between Systolic Blood Pressure and Death in Multivariate Subgroups Identified in At Least 20% of Trees Among Women From 3 Observational Cohort Studies (the Cardiovascular Health Study (1989–1993), the Health, Aging, and Body Composition Study (1997–1998), and the Sacramento Area Latino Study on Aging (1998–1999))

Subgroup	No. of Participants	%	Pooled Cohorts				NHANES				Consistent in Test Direction Sample	25% Difference From Total in Both Samples
			HR per 10-mm Hg Increase in SBP	95% CI	P Value	HR per 10-mm Hg Increase in SBP	95% CI	P Value				
Overall	5,931		1.11	1.10, 1.13	<0.001	1.11	1.07, 1.15	<0.001				
Creatinine >0.7 mg/dL												
Age >70 years	2,536	43	1.10	1.08, 1.13	<0.001	1.06	1.00, 1.13	0.04	Yes			
Age ≤70 years	1,287	22	1.14	1.09, 1.19	<0.001	1.06	0.93, 1.20	0.38				Yes
Creatinine ≤0.7 mg/dL												
No hypertension	1,080	18	1.12	1.05, 1.19	0.001	1.16	1.07, 1.26	<0.001	Yes			
Hypertension	937	16	1.00	0.96, 1.03	0.97	1.09	0.99, 1.20	0.07	Yes			
LDL cholesterol >130 mg/dL												
Creatinine >0.7 mg/dL	1,875	32	1.12	1.09, 1.15	<0.001	1.23	1.09, 1.38	0.001	Yes			
Creatinine ≤0.7 mg/dL	912	15	1.03	0.99, 1.07	0.2	1.18	1.03, 1.35	0.02				Yes
LDL cholesterol ≤130 mg/dL												
3MS score >80	2,650	45	1.15	1.12, 1.18	<0.001	1.14	1.00, 1.29	0.06	Yes			Yes
3MS score ≤80	350	6	1.04	0.98, 1.11	0.17	0.94	0.77, 1.15	0.52	Yes			Yes

Abbreviations: CI, confidence interval; HR, hazard ratio; LDL, low-density lipoprotein; 3MS, Modified Mini-Mental State Examination; NHANES, National Health and Nutrition Examination Survey; SBP, systolic blood pressure.

Table 3. Association Between Systolic Blood Pressure and Death in Multivariate Subgroups Identified in At Least 20% of Trees Among Men From 3 Observational Cohort Studies (the Cardiovascular Health Study (1989–1993), the Health, Aging, and Body Composition Study (1997–1998), and the Sacramento Area Latino Study on Aging (1998–1999))

Subgroup	No. of Participants	%	Pooled Cohorts				NHANES				Consistent in Test Direction Sample	25% Difference From Total in Both Samples
			HR per 10-mm Hg Increase in SBP	95% CI	P Value	HR per 10-mm Hg Increase in SBP	95% CI	P Value				
Overall	4,668		1.05	1.03, 1.07	<0.001	1.09	1.05, 1.13	<0.001				
DBP >80 mm Hg												
Age >67 years	867	19	1.05	1.01, 1.10	0.009	1.20	1.06, 1.34	0.003				
Age ≤67 years	305	7	1.25	1.13, 1.37	<0.001	1.14	0.98, 1.32	0.09	Yes	Yes	Yes	Yes
Body mass index ^a >30												
DBP >80 mm Hg	321	7	1.11	1.01, 1.20	0.02	1.14	0.94, 1.38	0.18	Yes	Yes	Yes	Yes
DBP ≤80 mm Hg	646	14	1.01	0.95, 1.08	0.71	1.03	0.92, 1.16	0.56	Yes	Yes	Yes	Yes
Not a fast walker												
No CHD	2,143	46	1.08	1.06, 1.11	<0.001	1.11	1.06, 1.16	<0.001	Yes	Yes	Yes	Yes
CHD	1,194	26	1.00	0.97, 1.04	0.79	1.08	0.98, 1.18	0.11	Yes	Yes	Yes	Yes

Abbreviations: DBP, blood pressure; CHD, coronary heart disease; CI, confidence interval; HR, hazard ratio; NHANES, National Health and Nutrition Examination Survey; SBP, systolic blood pressure.

^a Weight (kg)/height (m)².

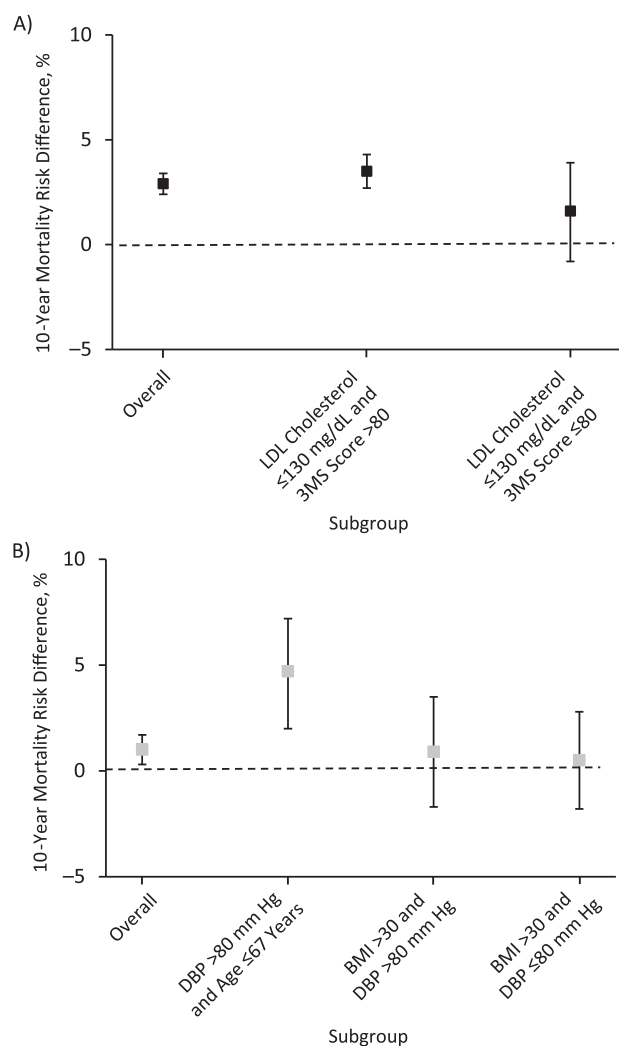


Figure 2. Ten-year mortality risk differences associated with a 10-mm Hg increase in systolic blood pressure among women (A) and men (B) from 3 pooled cohorts (the Cardiovascular Health Study (1989–1993), the Health, Aging, and Body Composition Study (1997–1998), and the Sacramento Area Latino Study on Aging (1998–1999)). Risk differences are presented overall and by subgroup. Body mass index (BMI) was calculated as weight (kg)/height (m)². Bars, 95% confidence intervals. DBP, diastolic blood pressure; LDL, low-density lipoprotein; 3MS, Modified Mini-Mental State Examination.

training set were stronger than the overall estimates among men with 1) DBP >80 mm Hg and age ≤67 years (hazard ratio (HR) = 1.25, 95% CI: 1.13, 1.37) and 2) BMI >30 and DBP >80 mm Hg (HR = 1.11, 95% CI: 1.01, 1.20). The subgroup-specific estimates for blood pressure were weaker than the overall estimates among men with BMI >30 and DBP ≤80 mm Hg (HR = 1.01, 95% CI: 0.95, 1.08).

We observed a single covariate combination in 20% of trees in the training set and did not include sex as a primary stratification variable. Among participants in the training set cohorts who had a gait speed less than 1.2 m/second, older age identified a subgroup in which higher SBP was

not associated with mortality. Among participants with gait speed <1.2 m/second, the hazard ratio for a 10-mm Hg increase in SBP was 1.03 (95% CI: 0.99, 1.06) among those aged >80 years and 1.08 (95% CI: 1.06, 1.09) among those aged ≤80 years. This pattern was replicated in NHANES: hazard ratios were 1.01 (95% CI: 0.96, 1.05) and 1.08 (1.04, 1.12), respectively.

In order to evaluate whether similar effect modification occurred on an additive scale, we examined the association of higher blood pressure with 10-year risk of mortality and observed similar patterning (Figure 2). Among women in the training set with LDL cholesterol ≤130 mg/dL and Modified Mini-Mental State Examination score >80, a 10-mm Hg increase in SBP was associated with a 3.5% higher risk of death over a 10-year period (95% CI: 2.7, 4.3%). The strongest estimate was again seen among men in the training set with DBP >80 mm Hg and age ≤67 years (risk difference = 4.7%, 95% CI: 2.0, 7.2).

DISCUSSION

Using data from observational studies, we identified HEAs between SBP and mortality in subpopulations. We identified 2 subpopulations in which higher blood pressure was consistently associated with a greater risk of mortality: women with low LDL cholesterol and preserved cognitive function and men aged 67 years or younger with high DBP. We also identified subgroups in which higher blood pressure was not associated with an increased risk of mortality, including adults aged 80 years or older with a slow gait. These groups may represent subpopulations of interest to evaluate in randomized controlled trials of blood pressure-lowering. The identification of HEAs can help generate hypotheses for a precision population health approach to improving blood pressure control.

Previous research has explored the role of heterogeneous treatment effects in the setting of blood pressure control. In a simulation study, Basu et al. (3) demonstrated that heterogeneous treatment effects could explain the apparently conflicting findings between 2 large randomized trials of blood pressure-lowering. Moreover, the top 3 winners of the *New England Journal of Medicine's* SPRINT Data Analysis Challenge aimed to better understand either individualized or subgroup (chronic kidney disease) decision-making regarding blood pressure control, based on consideration of the differential balance of benefits and harms in subpopulations (29). Baum et al. (4) used a random forest approach to identify heterogeneous treatment effects. They applied causal forest modeling to identify heterogeneous treatment effects in a trial of a weight loss intervention among people with type 2 diabetes (4). The authors found that hemoglobin A_{1c} levels and self-rated health could identify people with type 2 diabetes who were likely to benefit from the weight loss intervention.

We extended this previous work on heterogeneous treatment effects in trials to observational studies. A recent commentary in this journal highlighted the power of using observational studies to identify heterogeneity across populations (30). Observational studies often have larger and more representative study populations than trials; moreover,

not all biomarkers are amenable to intervention, but they could have differential associations in subpopulations. Our results are consistent with prior work that demonstrated that the association between blood pressure and outcomes differs across subgroups. Research has identified an attenuated or inverted relationship between blood pressure and outcomes in the very old, the frail, and those with disability, cognitive impairment, and low blood pressure (14, 15, 17, 31). Characteristics of the subpopulations in our study in which higher SBP was associated with mortality included being cognitively intact, young age, and elevated DBP. Low LDL cholesterol is the only characteristic we identified that had not been previously identified as an effect modifier of the association between SBP and mortality, although this could be a marker of statin use or adherence and related to cardiovascular risk.

There are several key considerations in the evaluation of HEAs. First is the decision of which variables should be examined as potential modifiers of exposure associations. We took a data-driven approach and included candidate variables that would potentially be measured in a clinical setting. An alternative approach would be to use a priori biological or clinical knowledge to select or preferentially weight variables for inclusion. Second, a concern about identification of HEAs is the scale used to measure the heterogeneity. We used the multiplicative scale through the use of hazard ratios to estimate our primary measure of effect, since this measure is commonly used in epidemiology and clinical medicine. However, it should be noted that effect measure modification is scale-specific (5, 27, 28). Since decision trees allow for the baseline risk to vary across each branch (subgroup), we repeated our analyses on the additive risk scale to ensure that the observed heterogeneity was not limited to the multiplicative scale. Third is the possibility of observing spurious differences among subgroups. Internal and external validation procedures are essential to minimize the likelihood of false-positive findings. We used a 2-stage validation procedure, considering only those subgroups that were identified in at least 20% of trees and then externally validating from this set of candidate trees on the test data. Notably, for 3 (21%) of the 14 groups identified in the internal sample, the exposure associations were not modified in the same direction in the external test sample. This highlights the heterogeneity of this population and the importance of external validation. A fourth consideration is whether the investigator wishes to limit the reporting of HEAs to differences of a certain magnitude. We defined an important difference as a difference in the β coefficient in the subgroup of interest and the overall sex-stratified coefficient of at least 25%. However, this threshold of 25% was an arbitrary cutpoint and depends on the unit of the exposure of interest (in this example, SBP was modeled per 10-mm Hg increment). Whether to include a similar restriction will depend on the goal of the analysis and what is identified as a meaningful difference.

Our goals in this paper were to further discussion of HEAs in the epidemiology literature and to provide an applied example in the setting of blood pressure. Identification of HEAs will be an essential component of precision medicine and could identify subgroups of interest for further inves-

tigation in randomized controlled trials or other studies. The identification of HEAs may help reconcile apparently conflicting results across studies, especially if the factors that modify the exposure associations are also associated with differing characteristics of the study participants. In this setting, reweighting methods can be used to adjust estimates to be more generalizable or to test whether the apparent difference between estimated effects is due to a systematic difference in the study population. Reweighting methods have been used to transport trial results to a specific target population (7). If investigators wish to make causal inferences from their findings, the assumptions of exchangeability, consistency, and positivity would need to be met; these assumptions are not addressed in this paper, but have been detailed elsewhere in the epidemiology literature (8–13).

However, some limitations of this approach should be considered. The primary limitation is that it can be logistically challenging to implement software code with which to identify HEAs using a random forest approach. We are developing an R package for this purpose and plan to share it with the scientific community. Others have used a similar approach to estimate heterogeneous treatment effects (32). A second limitation is that we examined a limited number of clinically measured variables to define the subgroups. There was observable and unobservable heterogeneity within the subgroups identified, and the estimated effect of an exposure in any given individual in a subgroup does not necessarily reflect the risk associated with that factor in the subgroup as a whole (5). Additionally, the nature of grouping in and of itself assumes homogenous effects within a subsample. A third limitation is that we assumed a linear association between SBP and the risk of death within subgroups; this decision was made to simplify the model, but a more sophisticated approach would allow for nonlinear relationships.

In summary, HEAs can help advance sciences toward a better understanding of how exposures act in complex subgroups. HEAs can guide the design of randomized controlled trials to those groups in which there may be uncertain benefit of exposure control, and they have the potential to help us better understand the pathophysiology of disease in diverse populations. Additional tools that help scientists validly and reliably identify HEAs are needed to advance the field of precision medicine.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology and Population Health, School of Medicine, Stanford University, Stanford, California (Michelle C. Odden); School of Biological and Population Health Sciences, Oregon State University, Corvallis, Oregon (Michelle C. Odden and Andreea M. Rawlings); School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, Oregon (Abtin Khodadadi and Xiaoli Fern); Department of Medicine, San Francisco VA Medical Center, San Francisco, California (Michael G. Shlipak, Kenneth Covinsky, Carmen A. Peralta); Kidney Health

Research Collaborative, University of California, San Francisco, San Francisco, California (Michael G. Shlipak and Carmen A. Peralta); Department of Medicine, School of Medicine, University of California, San Francisco, San Francisco, California (Michael G. Shlipak, Kirsten Bibbins-Domingo, Kenneth Covinsky, Alka Kanaya, Carmen A. Peralta); Department of Epidemiology and Biostatistics, School of Medicine, University of California, San Francisco, San Francisco, California (Michael G. Shlipak, Kirstin Bibbins-Domingo, Anne Lee, Mary N. Haan); Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania (Anne B. Newman); Cardiovascular Health Research Unit, Department of Medicine, School of Medicine, University of Washington, Seattle, Washington (Bruce M. Psaty); Departments of Epidemiology and Health Services, School of Public Health, University of Washington, Seattle, Washington (Bruce M. Psaty); and Kaiser Permanente Washington Health Research Institute, Seattle, Washington (Bruce M. Psaty).

This research was supported by National Institute on Aging (NIA) grant R01AG46206. The Cardiovascular Health Study was supported by contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, and N01HC85086 and grants U01HL080295 and U01HL130114 from the National Heart, Lung, and Blood Institute, with an additional contribution from the National Institute of Neurological Disorders and Stroke. Additional support was provided by NIA grant R01AG023629. The Health, Aging, and Body Composition Study was supported by NIA contracts N01-AG-6-2101, N01-AG-6-2103, and N01-AG-6-2106, NIA grant R01-AG028050, and National Institute of Nursing Research grant R01-NR012459, and in part by the Intramural Research Program of the National Institutes of Health, NIA. The Sacramento Area Latino Study on Aging was supported by NIA grant R01AG12975.

A full list of Cardiovascular Health Study principal investigators and institutions can be found at CHS-NHLBI.org.

The opinions expressed in this article are our own and do not reflect the views of the National Institutes of Health, the US Department of Health and Human Services, or the United States government.

Conflict of interest: none declared.

REFERENCES

1. Kent DM, Rothwell PM, Ioannidis JP, et al. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. *Trials*. 2010;11(1):Article 85.
2. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med*. 2015;372(9):793–795.
3. Basu S, Sussman JB, Hayward RA. Detecting heterogeneous treatment effects to guide personalized blood pressure treatment: a modeling study of randomized clinical trials. *Ann Intern Med*. 2017;166(5):354–360.
4. Baum A, Scarpa J, Bruzelius E, et al. Targeting weight loss interventions to reduce cardiovascular complications of type 2 diabetes: a machine learning-based post-hoc analysis of heterogeneous treatment effects in the Look AHEAD Trial. *Lancet Diabetes Endocrinol*. 2017;5(10):808–815.
5. Dahabreh IJ, Hayward R, Kent DM. Using group data to treat individuals: understanding heterogeneous treatment effects in the age of precision medicine and patient-centred evidence. *Int J Epidemiol*. 2016;45(6):2184–2193.
6. VanderWeele TJ. On the distinction between interaction and effect modification. *Epidemiology*. 2009;20(6):863–871.
7. Westreich D, Edwards JK, Lesko CR, et al. Transportability of trial results using inverse odds of sampling weights. *Am J Epidemiol*. 2017;186(8):1010–1014.
8. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*. 2008;168(6):656–664.
9. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. *Int J Epidemiol*. 1986;15(3):413–419.
10. Mortimer KM, Neugebauer R, van der Laan M, et al. An application of model-fitting procedures for marginal structural models. *Am J Epidemiol*. 2005;162(4):382–388.
11. VanderWeele TJ. Concerning the consistency assumption in causal inference. *Epidemiology*. 2009;20(6):880–883.
12. Westreich D, Cole SR. Invited commentary: positivity in practice. *Am J Epidemiol*. 2010;171(6):674–677.
13. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11(5):550–560.
14. Odden MC, Peralta CA, Haan MN, et al. Rethinking the association of high blood pressure with mortality in elderly adults: the impact of frailty. *Arch Intern Med*. 2012;172(15):1162–1168.
15. Peralta CA, Katz R, Newman AB, et al. Systolic and diastolic blood pressure, incident cardiovascular events, and death in elderly persons: the role of functional limitation in the Cardiovascular Health Study. *Hypertension*. 2014;64(3):472–480.
16. Sabayan B, Oleksik AM, Maier AB, et al. High blood pressure and resilience to physical and cognitive decline in the oldest old: the Leiden 85-Plus Study. *J Am Geriatr Soc*. 2012;60(11):2014–2019.
17. Wu C, Smit E, Peralta CA, et al. Functional status modifies the association of blood pressure with death in elders: Health and Retirement Study. *J Am Geriatr Soc*. 2017;65(7):1482–1489.
18. James G, Witten D, Hastie T, et al. *An Introduction to Statistical Learning: With Applications in R*. New York, NY: Springer Publishing Company; 2013.
19. Fried LP, Borhani NO, Enright P, et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol*. 1991;1(3):263–276.
20. Newman AB, Haggerty CL, Goodpaster B, et al. Strength and muscle quality in a well-functioning cohort of older adults: the Health, Aging and Body Composition Study. *J Am Geriatr Soc*. 2003;51(3):323–330.
21. Haan MN, Mungas DM, Gonzalez HM, et al. Prevalence of dementia in older Latinos: the influence of type 2 diabetes mellitus, stroke and genetic factors. *J Am Geriatr Soc*. 2003;51(2):169–177.
22. The Systolic Hypertension in the Elderly Program (SHEP) Cooperative Research Group. Rationale and design of a

- randomized clinical trial on prevention of stroke in isolated systolic hypertension. *J Clin Epidemiol*. 1988;41(12): 1197–1208.
23. Ives DG, Fitzpatrick AL, Bild DE, et al. Surveillance and ascertainment of cardiovascular events. The Cardiovascular Health Study. *Ann Epidemiol*. 1995;5(4): 278–285.
 24. Colón-López V, Haan MN, Aiello AE, et al. The effect of age at migration on cardiovascular mortality among elderly Mexican immigrants. *Ann Epidemiol*. 2009;19(1):8–14.
 25. Breiman L. Random forests. *Mach Learn*. 2001;45(1): 5–32.
 26. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods*. 2018;15(4):233–234.
 27. Hernán MA. The hazards of hazard ratios. *Epidemiology*. 2010;21(1):13–15.
 28. Psaty BM, Koepsell TD, Manolio TA, et al. Risk ratios and risk differences in estimating the effect of risk factors for cardiovascular disease in the elderly. *J Clin Epidemiol*. 1990;43(9):961–970.
 29. *New England Journal of Medicine*; SPRINT Trial Investigators; National Heart, Lung, and Blood Institute. SPRINT Challenge Winning Submissions. 2017. <https://challenge.nejm.org/pages/winners>. Accessed January 30, 2018.
 30. Glymour MM, Bibbins-Domingo K. The future of observational epidemiology: improving data and design to align with population health. *Am J Epidemiol*. 2019;188(5):836–839.
 31. Messerli FH, Mancina G, Conti CR, et al. Dogma disputed: can aggressively lowering blood pressure in hypertensive patients with coronary artery disease be dangerous? *Ann Intern Med*. 2006;144(12):884–893.
 32. Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests. *J Am Stat Assoc*. 2018;113(523):1228–1242.