Supplemental Appendix

Machine-learning-based high-benefit approach versus conventional high-risk approach in blood pressure management

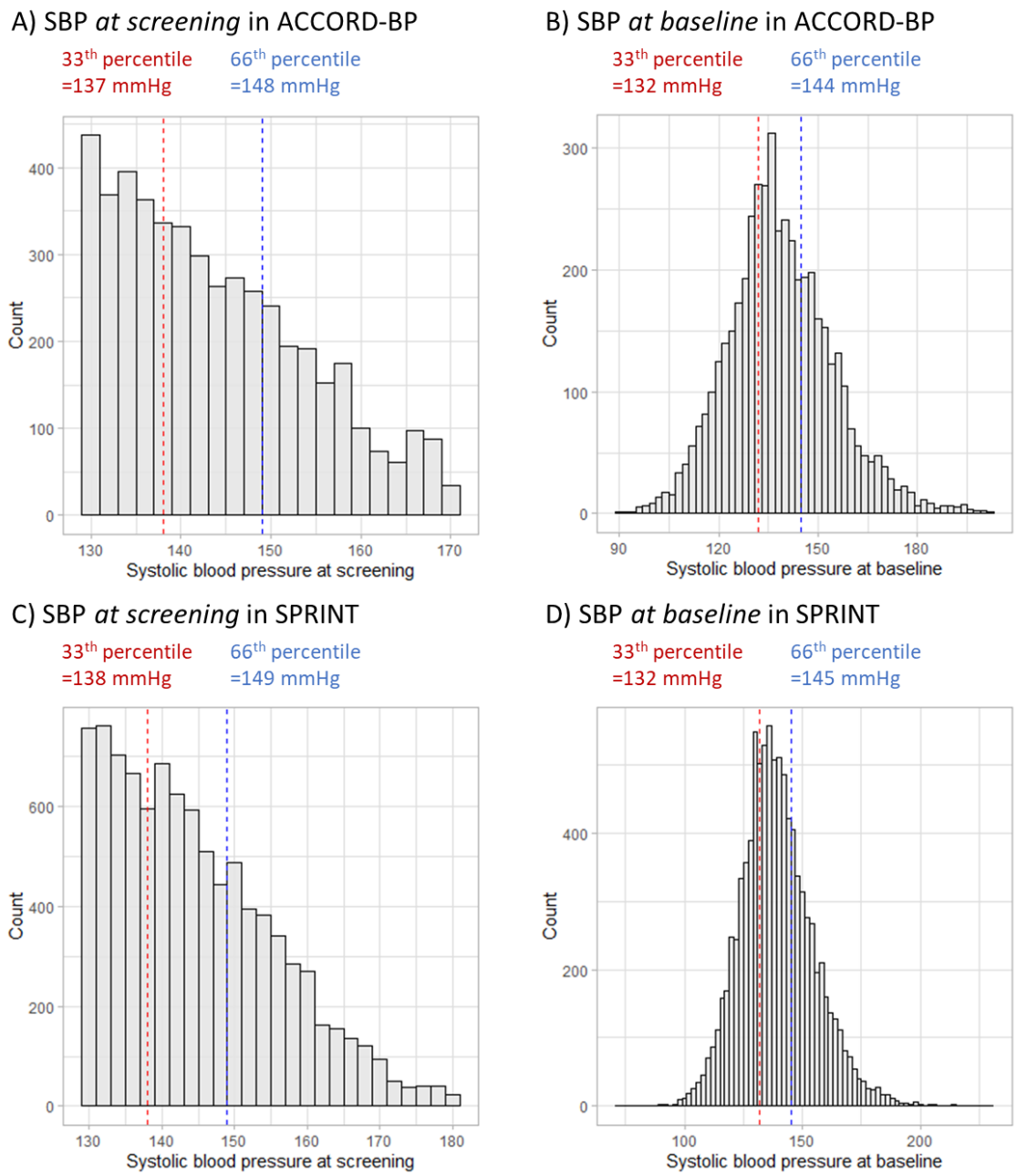
Supplementary Figures

1. Distribution of systolic blood pressure at screening and at baseline in the ACCORD-BP and the SPRINT trials.
2. Average treatment effects within each ranking as defined by predicted treatment effects using the causal forest method.
3. Distribution of reduction in cardiovascular outcomes by the ranking of predicted individual treatment effect (ITE) through the causal forest method.
4. Variable importance of the causal forest model.
5. Covariate balance plots between NHANES participants and SPRINT/ACCORD-BP trial samples before (blue) and after (yellow) inverse odds weighting to resemble NHANES participants in terms of covariates.

Supplementary Tables

1. Baseline characteristics of participants in the trial (SPRINT or ACCORD-BP) and the NHANES 1999-2018.
2. High-benefit approach vs High-risk approach using IPCW for right-censoring.
3. Comparing the performance of the high-benefit approach using our primary causal forest model for all bootstrap sample vs. building a causal forest model in each bootstrap sample.
4. High-benefit approach vs High-risk approach applying NHANES survey weights in transportability formula.
5. High-benefit approach vs High-risk approach based on 10-year ASCVD risk ≥20%

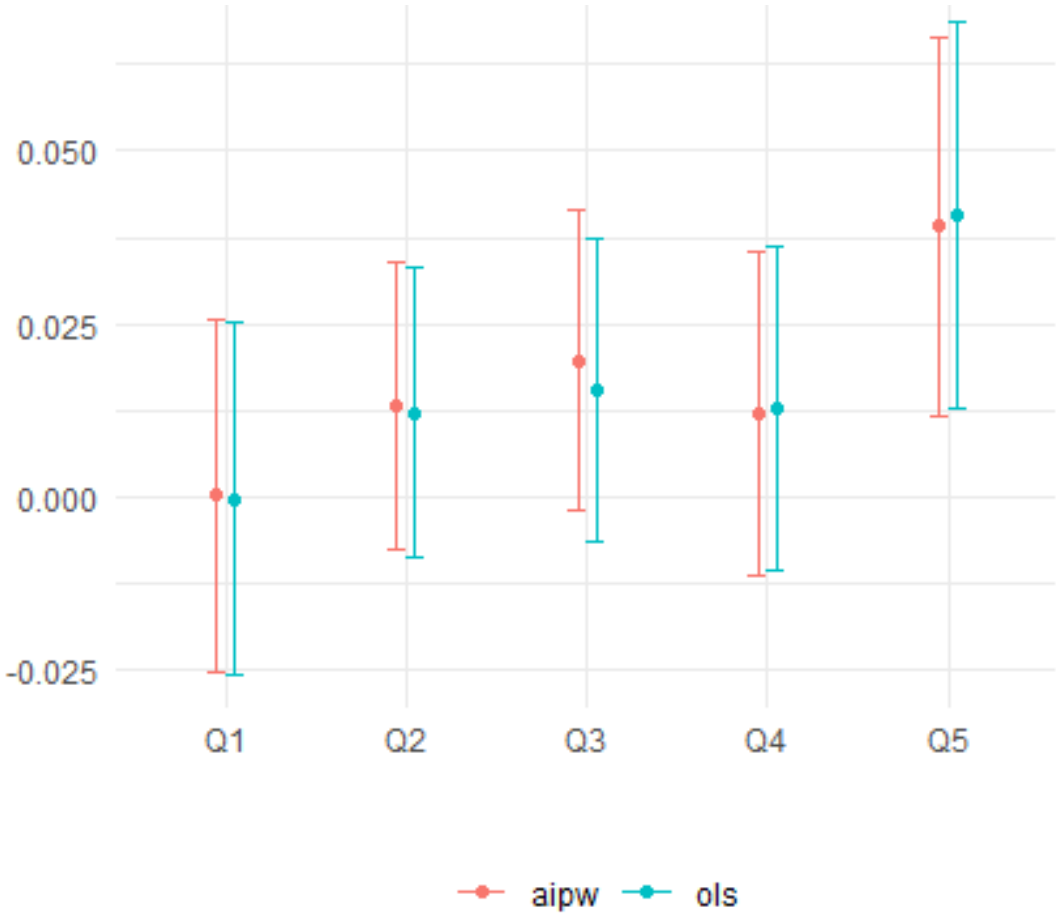
Supplementary Figure S1. Distribution of systolic blood pressure at screening and at baseline in the ACCORD-BP and the SPRINT trials.



SPRINT, Systolic Blood Pressure Intervention Trial; ACCORD-BP, Action to Control Cardiovascular Risk in Diabetes Blood Pressure.

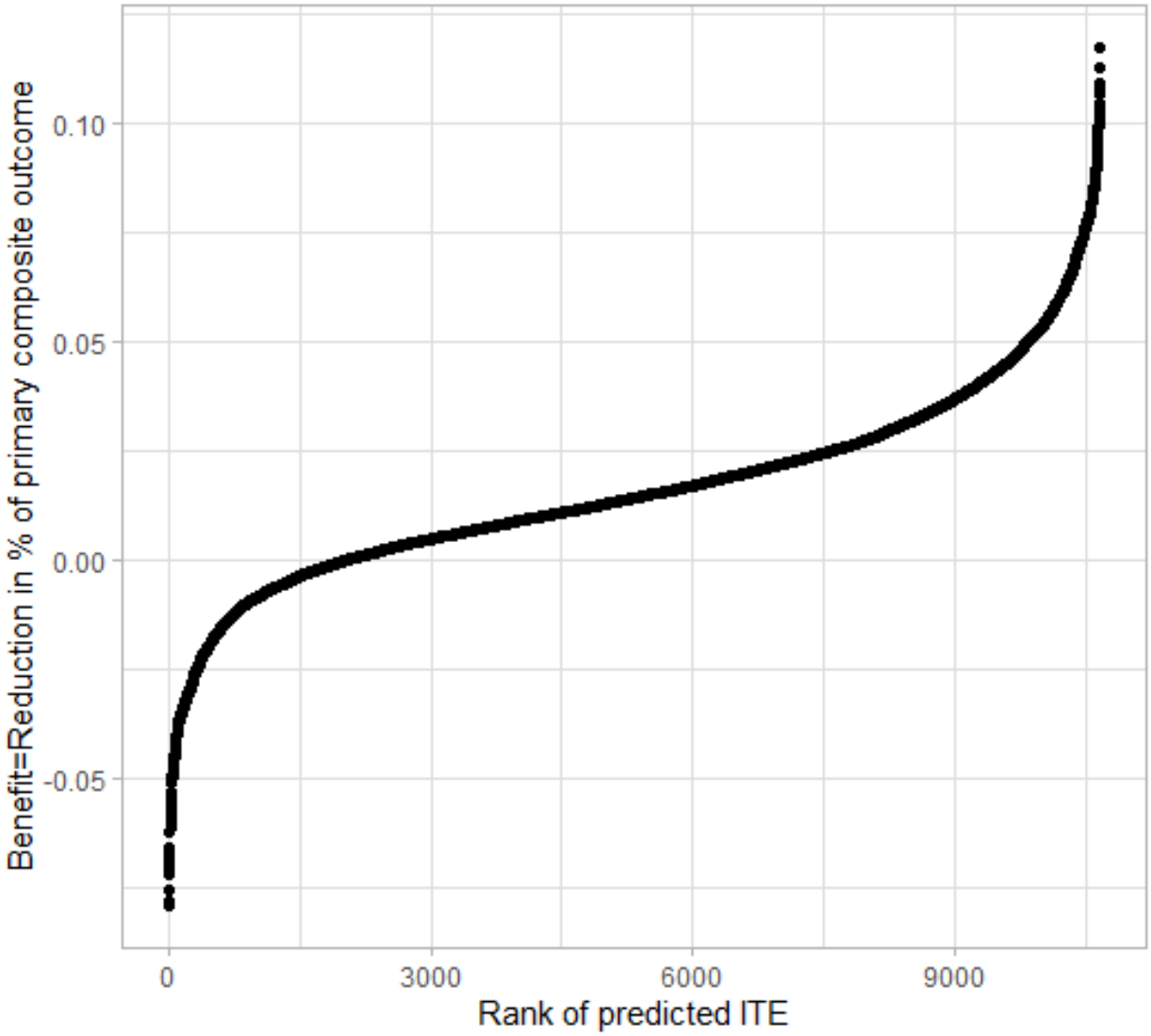
In both ACCORD-BP and SPRINT trials, participants were enrolled when they had an SBP of 130 to 180 mmHg *at screening*. However, *at baseline* (at the time of randomization), each individual showed different SBP levels from those *at screening.* For example, SPRINT protocol states that "It is not necessary for the randomization visit measurements to fall within eligibility criteria for the participant to be randomized" (ver 5.0. page 3b-3 "Collect all blood pressure related information"). We have categorized the participants into tertiles using SBP levels *at baseline* (rather than screening) to assess heterogeneity by SBP levels as the original articles of these trials used."

Supplementary Figure S2. Average treatment effects within each ranking as defined by predicted individual treatment effects using the causal forest method.

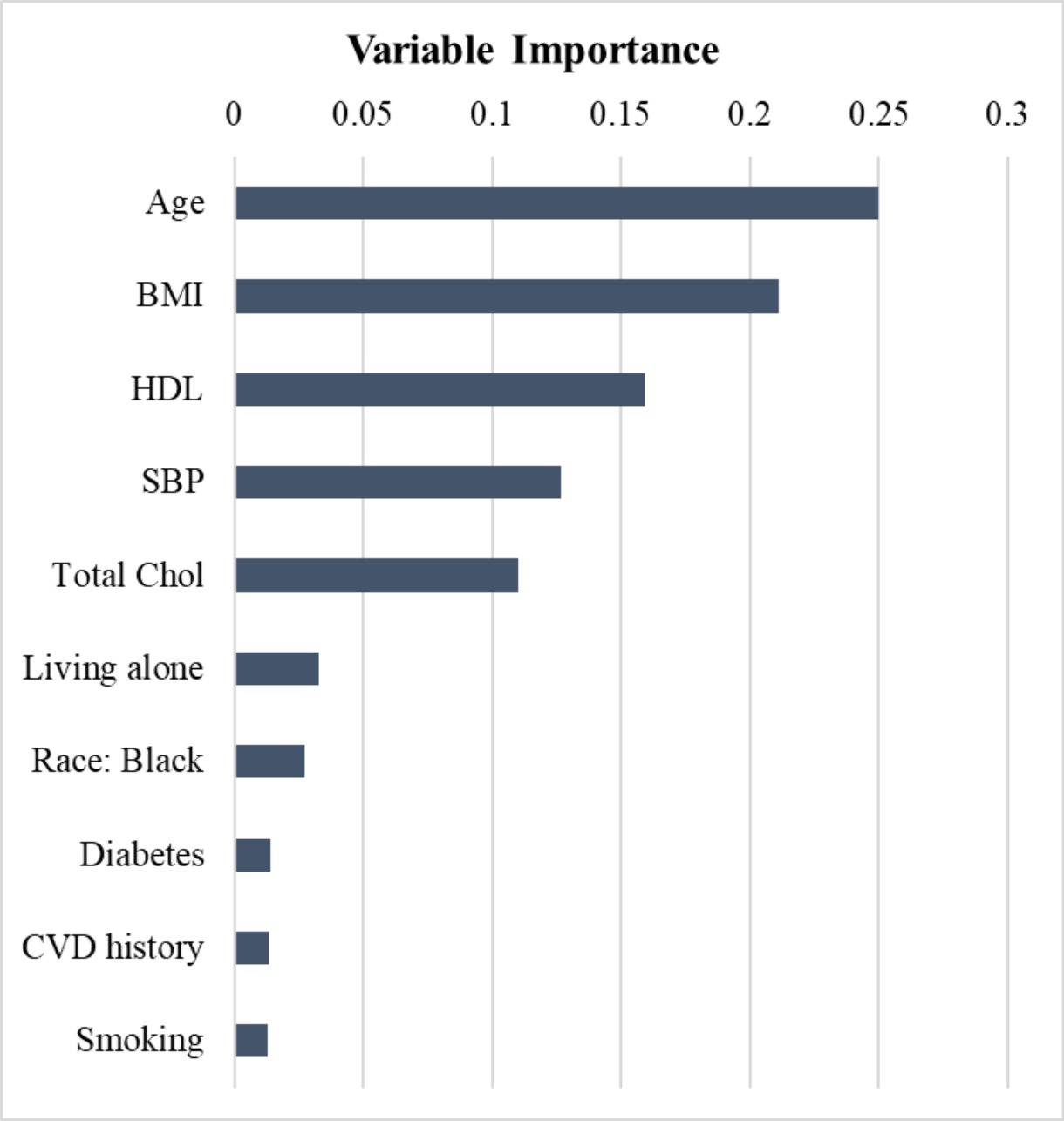


AIPW, augmented inverse propensity-weighted; OLS, ordinary least squares regression

Supplementary Figure S3. Distribution of reduction in cardiovascular outcomes by the ranking of predicted individual treatment effect (ITE) through the causal forest method.

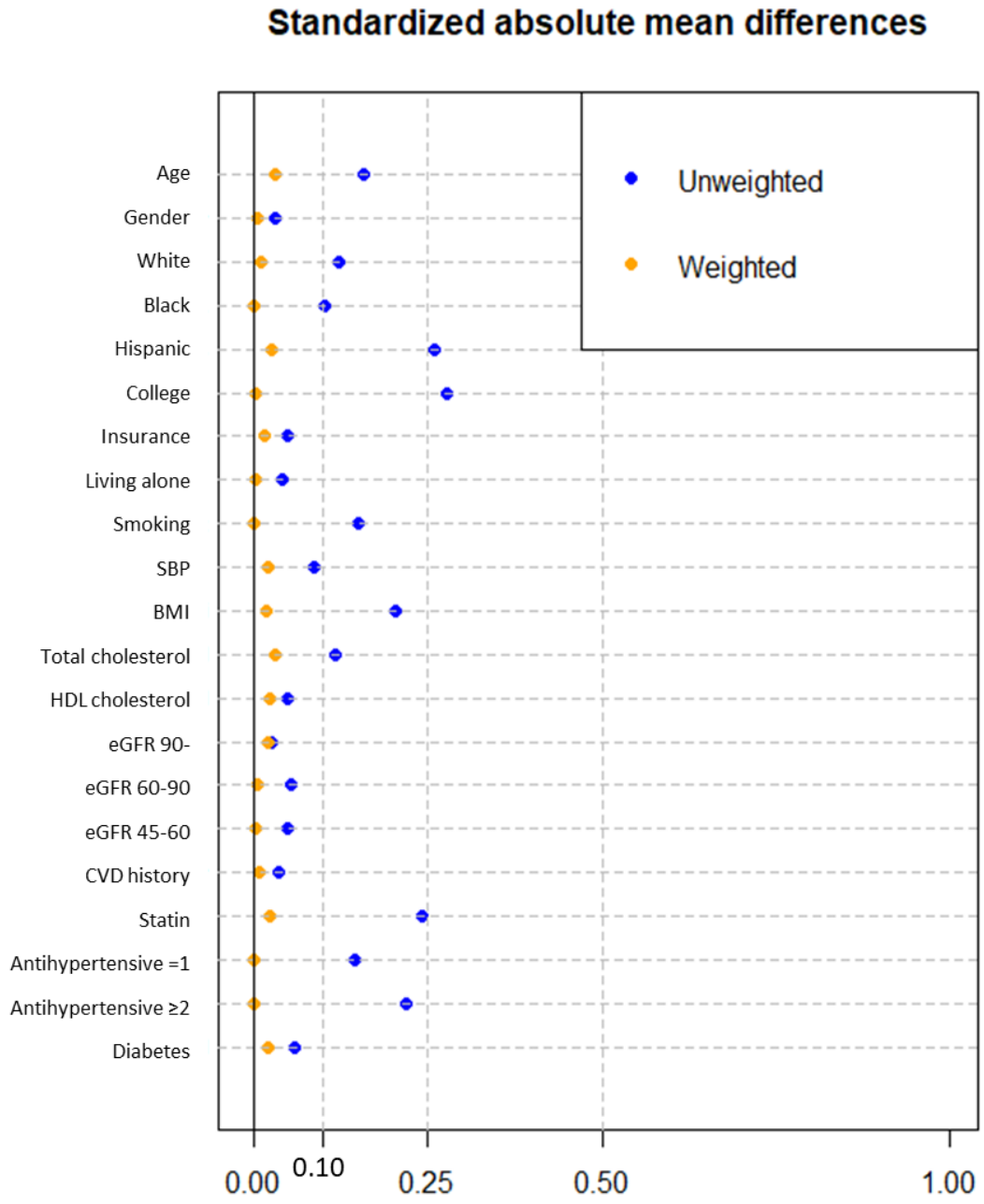


Supplementary Figure S4. Variable importance of the causal forest model.



BMI, body mass index; HDL, high-density lipoprotein; SBP, systolic blood pressure; CVD, cardiovascular disease
The variable importance was calculated by a simple weighted sum of how many times each variable was split at each depth in the causal forest. The top 10 variables are described in this Figure.

Supplementary Figure S5. Covariate balance plots between NHANES participants and SPRINT/ACCORD-BP trial samples before (blue) and after (yellow) inverse odds weighting to resemble NHANES participants in terms of covariates.



SPRINT, Systolic Blood Pressure Intervention Trial; ACCORD-BP, Action to Control Cardiovascular Risk in Diabetes Blood Pressure; NHANES, national health and nutrition examination survey; BMI, body mass index; SBP, systolic blood pressure; HDL, high-density lipoprotein; eGFR, estimated glomerular filtration rate; CVD, cardiovascular disease.

Supplementary Table S1. Baseline characteristics of participants in the trial (SPRINT or ACCORD-BP) and the NHANES 1999-2018.

Variables	**Trial participants in the SPRINT and ACCORD-BP (N=10,672)**	**NHANES 1999-2018 (N=14,575)**
Age, mean (SD), y	65.5 (8.4)	67.6 (10.3)
Female, %	4352 (40.8)	5631 (38.6)
Race/Ethnicity, %		
Non-Hispanic White	6267 (58.7)	7307 (50.1)
Non-Hispanic Black	2847 (26.7)	3002 (20.6)
Hispanic	668 (9.1)	3230 (22.2)
Others	890 (8.3)	1040 (7.1)
Education status, %		
Less than college	7130 (66.8)	12154 (83.4)
College or above	3542 (33.2)	2421 (16.6)
Uninsured, %	1339 (12.6)	1503 (10.3)
Living alone, %	2669 (25.3)	3332 (22.9)
Smoking, %	1395 (13.1)	3061 (21.0)
SBP, mean (SD), mmHg	139.3 (15.6)	137.0 (21.0)
BMI, mean (SD), kg/m^2	31.0 (5.3)	29.4 (6.2)
Total cholesterol, mean (SD), mg/dL	191.3 (42.5)	198.6 (45.6)
HDL cholesterol, mean (SD), mg/dL	49.8 (14.4)	50.9 (15.6)
eGFR, $mL/min/1.73\ m^2$, %		
≥90	3087 (28.8)	3987 (27.3)
60 to <90	5331 (50.0)	7054 (48.4)
45 to <60	1564 (14.7)	2333 (16.0)
<45	690 (6.5)	1210 (8.3)
Clinical CVD, %	2426 (22.7)	3000 (20.6)
Statin use, %	5653 (53.0)	5277 (36.2)
Antihypertensive use medications, %		
0	1113 (10.4)	5106 (35.0)
1	3586 (33.6)	3549 (24.4)
≥2	5973 (56.0)	5920 (40.6)
History of diabetes*	4483 (42.0)	5538 (38.0)
10-y Framingham CVD risk %, median (IQR)	22.7 (19.5)	25.5 (19.5)
10-year ASCVD risk %, median (IQR)	19.8 (18.0)	22.0 (20.4)

SPRINT, Systolic Blood Pressure Intervention Trial; ACCORD-BP, Action to Control Cardiovascular Risk in Diabetes Blood Pressure; NHANES, national health and nutrition examination survey; BMI, body mass index; SBP, systolic blood pressure; HDL, high-density lipoprotein; eGFR, estimated glomerular filtration rate; CVD, cardiovascular disease.

*History of diabetes was labeled as 0 for the SPRINT participants and 1 for the ACCORD-BP participants.

Supplementary Table S2. High-benefit approach vs High-risk approach using IPCW for right-censoring

Trial sample (N=10,672)	**High-benefit approach**	**High-risk approach #1 (Based on SBP)**	**High-risk approach #2 (Based on CVD risk score)**	**High-risk approach #3 (Based on CVD risk score)**
	Treat individuals with individualized treatment effect >0	Treat individuals with systolic blood pressure ≥130 mmHg	Treat individuals with 10-year Framingham CVD risk ≥20%	Treat individuals with 10-year ASCVD risk ≥10%
Sample average treatment effect (95% CI)	+10.47 pp (+9.30 to +11.61)	+2.49 pp (+1.17 to +3.85)	+3.08 pp (+1.47 to +4.87)	+2.76 pp (+1.49 to +4.04)
Difference (95% CI)	ref	+8.00 pp (+6.95 to +9.02)	+7.32 pp (+6.12 to +8.57)	+7.70 pp (+6.82 to +8.55)
Number needed to treat (95% CI)	10 (9 to 11)	40 (26 to 86)	32 (21 to 68)	36 (25 to 67)
Target population (N=14,575)	**High-benefit approach**	**High-risk approach #1 (Based on SBP)**	**High-risk approach #2 (Based on CVD risk score)**	**High-risk approach #3 (Based on CVD risk score)**
	Treat individuals with individualized treatment effect >0	Treat individuals with systolic blood pressure ≥130 mmHg	Treat individuals with 10-year Framingham CVD risk ≥20%	Treat individuals with 10-year ASCVD risk ≥10%
Population average treatment effect (95% CI)	+11.29 pp (+9.20 to +13.39)	+2.26 pp (+0.26 to +4.57)	+2.53 pp (-0.14 to +5.57)	+2.43 pp (+0.26 to +4.76)
Difference (95% CI)	ref	+8.99 pp (+7.14 to +10.87)	+8.69 pp (+7.02 to +10.59)	+8.84 pp (+7.55 to +4.76)
Number needed to treat (95% CI)	9 (7 to 11)	44 (22 to 386)	40 (18 to ∞)	41 (21 to 388)

IPCW, inverse-probability censoring weight; SBP, systolic blood pressure; CVD, cardiovascular disease; CI, confidence interval.
When we applied IPCW in our causal forest model, the coefficient of the mean forest prediction was 1.00 (p-value < 0.001) and the coefficient of the out-of-bag predicted treatment effect was 0.61 (p-value = 0.01) in the best linear fit model for the observed treatment effect.

Supplementary Table S3. Comparing the performance of the high-benefit approach using our primary causal forest model for all bootstrap samples vs. building a causal forest model in each bootstrap sample.

Trial sample (N=10,672)	**High-benefit approach (Standard approach of causal forest model)** [a]	**High-benefit approach (Modified approach of causal forest model)** [b]
Sample average treatment effect (95% CI)	+9.36 pp (+8.33 to +10.44)	+9.42 pp (+8.27 to +10.45)
Number needed to treat (95% CI)	11 (10 to 12)	11 (10 to 12)
Target population (N=14,575)	**High-benefit approach (Standard approach of causal forest model)** [a]	**High-benefit approach (Modified approach of causal forest model)** [b]
Population average treatment effect (95% CI)	+8.85 pp (+6.78 to +10.79)	+9.48 pp (+7.32 to +11.49)
Number needed to treat (95% CI)	11 (9 to 15)	11 (9 to 14)

CI, confidence interval.

[a]In the standard approach, we applied our primary causal forest model (with moderate to high calibration and discrimination performance) described in the main manuscript to estimate average treatment effect.

[b]In the modified approach, we build causal forest model in each of the bootstrap sample to estimate average treatment effect.

Supplementary Table S4. High-benefit approach vs High-risk approach applying NHANES survey weights in transportability formula.

Target population (N=14,575)	**High-benefit approach**	**High-risk approach #1 (Based on SBP)**	**High-risk approach #2 (Based on CVD risk score)**	**High-risk approach #3 (Based on CVD risk score)**
	Treat individuals with individualized treatment effect >0	Treat individuals with systolic blood pressure ≥130 mmHg	Treat individuals with 10-year Framingham CVD risk ≥20%	Treat individuals with 10-year ASCVD risk ≥10%
Population average treatment effect (95% CI)	+9.68 pp (+8.34 to +10.90)	+2.52 pp (+1.15 to +3.97)	+3.16 pp (+1.24 to +4.92)	+3.19 pp (+1.74 to +4.65)
Difference (95% CI)	ref	+7.18 pp (+6.07 to +8.21)	+6.53 pp (+5.16 to +7.96)	+6.49 pp (+5.41 to +7.55)
Number needed to treat (95% CI)	10 (9 to 12)	40 (26 to 84)	32 (21 to 70)	33 (23 to 63)

NHANES, national health and nutrition examination survey; SBP, systolic blood pressure; CVD, cardiovascular disease; CI, confidence interval.

Supplementary Table S5. High-benefit approach vs High-risk approach based on 10-year ASCVD risk ≥20%

A) Trial sample (N=10,672)	**High-benefit approach**	**High-risk approach #4 (Based on CVD risk score)**
	Treat individuals with individualized treatment effect >0	Treat individuals with 10-year ASCVD risk ≥20%
No. of individuals treated	8,563	5,266
Sample average treatment effect (95% CI)	+9.36 pp (+8.33 to +10.44)	+1.92 pp (+0.19 to +3.52)
Difference (95% CI)	ref	+7.44 pp (+6.22 to +8.82)
Number needed to treat (95% CI)	11 (10 to 12)	52 (28 to 526)
B) Target population (N=14,575)	**High-benefit approach**	**High-risk approach #4 (Based on CVD risk score)**
	Treat individuals with individualized treatment effect >0	Treat individuals with 10-year ASCVD risk ≥20%
No. of individuals treated	11,320	8,043
Population average treatment effect (95% CI)	+8.85 pp (+6.78 to +10.79)	+1.25 pp (-1.67 to +3.92)
Difference (95% CI)	ref	+7.60 pp (+5.93 to +9.32)
Number needed to treat (95% CI)	11 (9 to 15)	80 (26 to ∞)

Outcome was the reduction in % of primary composite CVD outcomes during a 3-year follow-up. The 95% confidence intervals (CIs) were calculated using 1000 bootstrapped samples. The average treatment effect and number needed to treat of each approach were obtained using the sample from the combined database of SPRINT and ACCORD-BP (trial sample), along with inverse-odds-weights to emulate the trial sample to the NHANES participants. Number needed to treat was calculated by 1/average treatment effect. The 10-year ASCVD risk for high-risk approach #4 was calculated by ACC/AHA pooled cohort equation.