



HHS Public Access

Author manuscript

Lancet Digit Health. Author manuscript; available in PMC 2022 December 21.

Published in final edited form as:

Lancet Digit Health. 2022 November ; 4(11): e796–e805. doi:10.1016/S2589-7500(22)00170-4.

Individualising intensive systolic blood pressure reduction in hypertension using computational trial phenomaps and machine learning: a post-hoc analysis of randomised clinical trials

Evangelos K Oikonomou,

Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA

Erica S Spatz,

Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA

Center for Outcomes Research and Evaluation, Yale–New Haven Hospital, New Haven, CT, USA

Marc A Suchard,

Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, Los Angeles, CA, USA

Department of Computational Medicine, David Geffen School of Medicine at UCLA, University of California, Los Angeles, Los Angeles CA, USA

Department of Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, Los Angeles CA, USA

Rohan Khera

Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA

This is an Open Access article under the [CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/) license.

Correspondence to: Dr Rohan Khera, Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT 06510, USA, rohan.khera@yale.edu.

Contributors

EKO contributed to study design, data analysis, statistical analysis, and writing of the manuscript. ESS and MAS provided scientific direction and wrote the report. RK designed the study, coordinated data analysis, directed the project, and wrote the report. EKO and RK accessed and verified the raw underlying data, and were responsible for the final decision to submit for publication. All authors had access to all the data in the study.

Declaration of interests

RK and EKO are co-founders of Evidence2Health, a machine learning health analytics company, from which they have received consultancy fees. RK and EKO are named as inventors in a US provisional patent application (titled “methods for neighborhood phenomapping for clinical trials”, number 63/177117, filed on April 20, 2021), related to the submitted work. RK receives grant support from the National Heart, Lung, and Blood Institute of the US National Institutes of Health, and the Doris Duke Charitable Foundation, outside of the submitted work. RK also receives grant support, through Yale University, from Bristol-Myers Squibb, and has served on the Bristol-Myers Squibb digital health advisory board, outside of the submitted work. EKO has received grant support for attending meetings from the American Heart Association (Council for Lifestyle and Cardiometabolic Health). MAS reports institutional grant support from the US National Institutes of Health, US Food and Drug Administration, and US Department of Veteran Affairs; personal consulting fees from Janssen Research and Development and Private Health Management; and institutional grant support from Advanced Micro Devices, outside the scope of the submitted work. All other authors declare no competing interests.

For the online browser-accessible version of PRECISION see <https://www.cards-lab.org/precision>

See Online for appendix

Center for Outcomes Research and Evaluation, Yale–New Haven Hospital, New Haven, CT, USA

Section of Health Informatics, Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA

Summary

Background—The cardiovascular benefits of intensive systolic blood pressure control vary across clinical populations tested in large randomised clinical trials. We aimed to evaluate the application of machine learning to clinical trials of patients without and with type 2 diabetes to define the personalised cardiovascular benefit of intensive control of systolic blood pressure.

Methods—In SPRINT, a trial of intensive (systolic blood pressure <120 mm Hg) versus standard (systolic blood pressure <140 mm Hg) systolic blood pressure control in patients without type 2 diabetes, we defined a phenotypic representation of the study population using 59 baseline variables. We extracted personalised treatment effect estimates for the primary outcome, time-to-first major adverse cardiovascular event (MACE; cardiovascular death, myocardial infarction or acute coronary syndrome, stroke, and acute decompensated heart failure), through iterative Cox regression analyses providing average hazard ratio (HR) estimates weighted for the phenotypic distance of each participant from the index patient of each iteration. Next, we trained an extreme gradient boosting algorithm (known as XGBoost) to predict the personalised effect of intensive systolic blood pressure control using features most consistently linked to increased personalised benefit, before evaluating its performance in the ACCORD BP trial of patients with type 2 diabetes randomly assigned to receive intensive versus standard systolic blood pressure control. We stratified patients based on their predicted treatment effect, and key demographic groups (age, sex, cardiovascular disease, and smoking). We assessed the presence of heterogeneity with an interaction test, and assessed the performance of the algorithm in a simulation analysis of SPRINT in the presence or absence of an artificially introduced heterogeneous treatment effect.

Findings—From SPRINT, we included all 9361 study participants (mean age 67.9 years [SD 9.4], 3332 [35.6%] female) who underwent randomisation to either intensive (n=4678) or standard (n=4683) treatment. The median individualised HR for MACE was 0.63 (IQR 0.53–0.78). An eight-feature tool built for this analysis to predict personalised benefit in SPRINT was externally tested in ACCORD BP (4733 participants (mean age 62.7 years [SD 6.7], 2258 [47.7%] female), wherein it successfully identified individuals with differential benefit from intensive versus standard systolic blood pressure control (adjusted HR for MACE of 0.70 [95% CI 0.55–0.90] in individuals with above-median MACE benefit versus 1.05 [95% CI 0.84–1.32] for below-median predicted benefit; $p_{\text{interaction}}=0.0184$). Subgroup analysis based on age (<65 years: HR 0.89 [95% CI 0.71–1.12]; ≥65 years: 0.85 [0.67–1.09]), sex (male: 0.89 [0.72–1.10]; female: 0.85 [0.65–1.10]), established cardiovascular disease (no: 0.89 [0.70–1.14]; yes: 0.84 [0.67–1.06]), or active smoking (no: 0.85 [0.71–1.02]; yes: 1.01 [0.64–1.60]) did not identify groups with heterogeneity of treatment effect. In a simulation analysis of SPRINT, the proposed algorithm detected groups with heterogeneous treatment effects in the presence, but not absence, of simulated subgroup differences.

Interpretation—By use of machine learning to define an individual's personalised benefit through phenotypic representations of clinical trials, we created a practical tool for individualising

the selection of intensive versus standard systolic blood pressure control in patients without and with type 2 diabetes.

Funding—National Heart, Lung, and Blood Institute of the US National Institutes of Health.

Introduction

Hypertension is the modifiable metabolic risk factor with the largest contribution to cardiovascular disease burden globally.¹ Isolated systolic hypertension is the most common form of hypertension,² with systolic blood pressure considered a stronger cardiovascular risk factor than diastolic blood pressure.³ The Systolic Blood Pressure Intervention Trial (SPRINT)⁴ and Action to Control Cardiovascular Risk in Diabetes Blood Pressure trial (ACCORD BP)⁵ tested whether targeting a systolic blood pressure of less than 120 mm Hg (intensive group) reduces the incidence of a major adverse cardiovascular event (MACE) when compared with a systolic blood pressure target of less than 140 mm Hg (standard group). Although both trials included individuals with a systolic blood pressure of 130–180 mm Hg who had or were at an elevated risk of cardiovascular disease, type 2 diabetes was an exclusion criterion in SPRINT and an eligibility requirement in ACCORD BP. Intensive systolic blood pressure treatment demonstrated a substantial cardiovascular benefit in SPRINT,^{4,6} but similar benefits were not seen in ACCORD BP.^{5,7}

These discordant findings might suggest that the benefits of intensive systolic blood pressure control depend on the phenotypic profile of each patient. Indeed, clinical practice guidelines have suggested that systolic blood pressure targets should be based on a patient's cardiovascular risk profile;⁸ however, risk profiles alone do not adequately capture the phenotypic diversity of individuals, which is a requirement for precision care.

Our research group has recently published a machine learning-based approach that enables phenotyping of a clinical trial population based on baseline characteristics to predict treatment response.^{9,10} By defining a computational trial phenomap, a mathematical construct of the individual phenotypes across all these baseline measures, we can evaluate heterogeneous treatment effects via computational approaches that account for the phenotypic similarity of all included individuals, the treatment each individual received, and their subsequent clinical outcomes.

In this study, we aimed to test our hypothesis that patients with hypertension exhibit differential cardiovascular benefits from intensive versus standard systolic blood pressure reduction based on their complex phenotypic profile at baseline. Using participant-level data from SPRINT, we developed a computational phenomapping strategy that leverages information from all trial participants to infer signatures of individualised benefit of intensive systolic blood pressure lowering among patients without type 2 diabetes.^{4,6} We then evaluated its ability to identify patients with type 2 diabetes in the ACCORD BP trial who benefitted from intensive systolic blood pressure control.⁵

Methods

Data sources

We obtained participant-level data of the SPRINT (n=9361) and ACCORD BP (n=4733) trials through the National Heart, Lung, and Blood Institute Biologic Specimen and Data Repository Information Coordinating Center. The design and original results for both studies have been published.⁴⁻⁶ A detailed description of the patient population is provided in the appendix (p 1). In both the SPRINT and ACCORD BP trials, participants were randomly assigned (1:1) to a systolic blood pressure goal of either less than 120 mm Hg (intensive treatment) or less than 140 mm Hg (standard treatment). A glossary of terms can be found in the appendix (p 10).

This post-hoc analysis of de-identified data complied with the Declaration of Helsinki and was approved by the Yale Institutional Review Board, which waived the requirement for informed consent.

Study population and covariates

We included all SPRINT and ACCORD BP participants who underwent randomisation and were included in the original trial reports.⁴⁻⁶ Inclusion and exclusion criteria for these trials have been described previously.⁴⁻⁶ No patients were further excluded for this analysis. In SPRINT, which was used as the derivation trial, we included patient characteristics at trial enrolment and the index or baseline visit, including demographics, anthropometric indices, medication, cardiovascular history, non-cardiovascular history, physical activity, self-reported assessment of health status, blood pressure measurements, Framingham Risk Score of 10-year atherosclerotic cardiovascular disease risk, laboratory and electrocardiographic measurements, and cognitive testing results. All features are listed in the appendix (pp 11–12).

Data pre-processing

The data pre-processing steps are described in detail in the appendix (pp 1–2). Briefly, we reviewed all clinically relevant variables collected in SPRINT (10% missingness), before imputing any missing information using chained random forests with predictive mean matching (*missRanger* in R).¹¹ Following normalisation, reduction of collinearity, and removal of near-zero variance factors, 59 variables were included in our analysis (appendix pp 11–12).

Outcomes

Consistent with SPRINT,^{4,6} our primary outcome was the incidence of the first MACE, defined as a composite of myocardial infarction or acute coronary syndrome, stroke, acute decompensated heart failure, or cardiovascular mortality.

Our secondary outcome was the incidence of a composite net clinical benefit endpoint, including the primary outcome, all-cause mortality, and serious adverse events. Serious adverse events were defined as events deemed to be life threatening, leading to

hospitalisation or prolonging hospitalisation, or resulting in persistent or substantial disability or death. All analyses were performed in an intention-to-treat manner.

Defining a computational trial phenomap

We computed phenotypic distances between individuals based on their baseline characteristics according to the Gower's distance, which is a metric of dissimilarity between two patients based on mixed continuous and categorical data (appendix p 2).¹² To visualise the phenotypic variation in the SPRINT population, we used uniform manifold approximation and projection (appendix pp 2–3, 10). This method constructs a phenomap (two-dimensional representation) of the population based on the full breadth of baseline phenotypes seen in the trial.¹³ This projection allows a degree of interpretability of the distribution of patients in this multidimensional phenotypic space, through colour-coded maps.

Defining personalised treatment effect estimates

For each unique participant, we measured the association of intensive versus standard systolic blood pressure reduction with the primary and secondary outcomes of interest using weighted estimation by Cox regression, as proposed by Schemper and colleagues.¹⁴ Weights for our analysis were derived from the dissimilarity matrix with the contribution of each patient to the final prediction dependent on their phenotypic distance from the index patient. To ensure that patients who are phenotypically closer to the index patient carried higher cumulative weights than did patients located further away, we evaluated kernels with different exponential transformations of the similarity metric, defined as (1–Gower's distance). These values were processed through a rectified linear unit function (with SoftMax pre-processing) before their inclusion as weights in the regression models. We also analysed discrete phenotypic neighbourhood sizes consisting of 5% or 10% of the phenotypically similar patients within each patient's neighbourhood, as per our previous work.^{9,10} We selected the weight definition that provided the most consistent estimates with neighbourhood-based methods (appendix pp 3–5, 14). From each personalised Cox regression model, we extracted the natural logarithmic transformation of the hazard ratio (log HR) comparing intensive versus standard systolic blood pressure control. Negative values favoured intensive systolic blood pressure control as more protective against the outcome of interest. Furthermore, for each patient, we calculated the difference in the weighted mean systolic blood pressure between the intensive and standard treatment groups at 1, 6, and 12 months.

Developing and validating an algorithm to identify benefit from intensive systolic blood pressure control

We trained an extreme gradient-boosting algorithm (XGBoost)¹⁵ to predict the personalised log HRs of MACE with intensive versus standard systolic blood pressure reduction using a subset of 32 baseline variables (appendix pp 11–12), which were selected based on their availability and consistent definitions between SPRINT and ACCORD BP. The model was trained in a randomly selected subset of SPRINT participants, consisting of 80% of the study population, with internal validation done in the remaining 20% of the trial population. Random sampling was done in R.

Specifically, the XGBoost algorithm was constructed to identify the baseline phenotypes most strongly linked to the individualised log HR values in a Cox regression model. Model performance was evaluated using root mean square error as the loss function.⁹ Optimal hyperparameters, including the learning rate and those defining the depth and structure of the tree-based architecture, were selected (appendix pp 5–6) using a random search, and five-fold cross-validation for internal validation. To allow for interpretation of our model's predictions, we assessed feature importance using Shapley additive explanations (SHAP) values to identify a predictor's relative contribution, either positively or negatively, to the final prediction.¹⁶ Further details are given in the appendix (pp 6–7).

To improve the model's practical application, we selected features that were most strongly associated with the cardiovascular effects of an intensive treatment strategy based on a SHAP feature importance of 0.01 or higher, identifying eight features. These features were sex, renal function (glomerular filtration rate or creatinine), history of coronary disease requiring revascularisation, history of angina, active smoking, and statin or aspirin use. We retrained our model using this set of features, based on the same approach described above, and arrived at a parsimonious tool to predict the personalised benefit of targeting a systolic blood pressure goal of less than 120 mm Hg versus less than 140 mm Hg. The final tool was named PRECISION (for PREssure Control In hypertenSION) and an online browser-accessible version of PRECISION was also made available for external use.

Performance of PRECISION in ACCORD BP

We evaluated PRECISION in the external ACCORD BP trial to provide patient-level predictions on the expected cardiovascular effect of intensive versus standard systolic blood pressure treatment among patients with type 2 diabetes. We examined the relative hazard of MACE across strata of predicted response (predicted benefit above versus below the median predicted response) and actual group randomisation (intensive versus standard systolic blood pressure control), with an interaction test used to assess heterogeneity between subgroups. Subgroups were age, sex, established cardiovascular disease, and active smoking.

Sensitivity analysis through simulation studies

To assess the sensitivity of the algorithm in detecting treatment effect heterogeneity in a clinical trial population, we performed a positive and negative control simulation study. Full details of the simulation are given in the appendix (pp 8–9). Briefly, using the baseline SPRINT characteristics, we introduced artificial endpoints in the presence or absence of treatment effect heterogeneity in a predefined subgroup (ie, women on aspirin [n=1470, 15.7% of entire cohort] versus the rest of the cohort). Treatment effect heterogeneity was introduced using the method proposed by Rigdon and colleagues,¹⁷ with our first simulation (positive control) introducing an average treatment effect of 6% among women with aspirin versus no effect among the rest of the cohort, and the second simulation (negative control) introducing an average treatment effect of 1% for all groups. Follow-up time data were introduced at random using a Gompertz distribution. Similar to our analysis in the original SPRINT dataset, we performed phenomapping of the baseline trial population and extracted individualised treatment effect estimates. Following this, we trained an XGBoost model in the randomly selected training set (n=6242, 67%), before applying the algorithm in the

remaining (test) set (n=3119, 33%) to (1) compare the predicted effect estimates between our predefined patient subgroups (women on aspirin versus the rest of the cohort); and (2) allow the algorithm to identify de novo subgroups with treatment effect heterogeneity.

Statistical analysis

Categorical variables are summarised as numbers (percentages), and continuous variables as mean values with standard deviation or median with IQR (Q1–Q3), as appropriate. Continuous variables between two groups were compared using Student's *t* test. Pearson's *r* was used to assess the pairwise correlation between continuous variables. When extracting patient-specific effect estimates during the training stage, we applied weighted Cox estimation as proposed by Schemper and colleagues (implemented in the R package *coxphw*).¹⁴ This method enables unbiased average HR estimates in case of non-proportional hazards (appendix pp 3–5). In sensitivity analyses, we assessed the correlation between the individualised log(HR) (relative risk reduction) and (1) the cumulative hazard of the primary outcome in the control group and (2) the observed absolute risk reduction, both at the median follow-up of 3.26 years based on Kaplan-Meier analyses in SPRINT weighted for each participant using the previously defined weights.⁴ We explicitly adjusted our Cox regression models for age and sex given their importance in defining clinical patient groups. We performed survival regression analyses using time-to-first event Cox regression models, which are graphically presented as unadjusted Nelson-Aalen plots. We assessed between-subgroup heterogeneity using a test for interaction.

In a post-hoc evaluation of the model in ACCORD BP, we also applied an inverse probability censoring weighted estimator, evaluating whether features identified by our algorithm as defining heterogeneous treatment effects were robust to potential effects of dependent censoring, using the method proposed by Willems and colleagues (appendix pp 7–8).¹⁸

All statistical tests were two-sided with a level of significance of 0.05 without correction for multiplicity of comparisons. Analyses were performed using R (version 4.0.2) and Python (version 3.8.5). Reporting of the study design and findings is in accordance with the STROBE guidelines.¹⁹

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the manuscript.

Results

From SPRINT, we included all 9361 study participants (mean age 67.9 years [SD 9.4], 3332 [35.6%] female, 5399 [57.7%] non-Hispanic White, 2802 [29.9%] non-Hispanic Black, and 984 [10.5%] Hispanic) who underwent randomisation to either intensive (n=4678) or standard (n=4683) treatment. Participants were followed up over a mean period of 3.8 years (SD 1.0), during which a first primary MACE event was reported in 562 individuals. For the secondary outcome of net clinical benefit, there were 365 total death events and 3529 serious adverse events, with the outcome occurring in 3549 participants. In ACCORD BP,

4733 participants (mean age 62.7 years [SD 6.7], 2258 [47.7%] female, 2864 [60.5%] non-Hispanic White, 1142 [24.1%] Black, and 330 [7.0%] Hispanic) underwent randomisation to either intensive (n=2362) or standard (n=2371) treatment. Participants were followed up for a mean period of 4.9 years (SD 1.2) with 553 first MACE events recorded. Secondary outcomes were not available in a manner that enabled a direct comparison with SPRINT.

The phenomap of the SPRINT trial was based on pairwise distances between all trial participants according to the Gower's dissimilarity index using 59 baseline phenotypic variables (figure 1). Visual assessment of the risk phenomaps showed that the treatment groups were randomly distributed in the phenomic space (figure 1A). By contrast, baseline phenotypic variables, such as sex, age, 10-year Framingham Risk Score for atherosclerotic cardiovascular disease, and systolic and diastolic blood pressure were heterogeneously distributed reflecting distinct phenotypic neighbourhoods (figure 1B–F). Longitudinal blood pressure monitoring confirmed a greater weighted mean systolic blood pressure reduction in the intensive group than in the standard group, over time, with overall consistent effects across phenotypic neighbourhoods (median neighbourhood difference of -5.4 mm Hg [IQR -6.0 to -4.9] at 1 month; -13.0 mm Hg [-13.6 to -12.5] at 6 months; and -15.4 mm Hg [-15.9 to -15.0] at 12 months; appendix p 15).

We subsequently calculated individualised estimates of cardiovascular and net clinical benefit with intensive versus standard systolic blood pressure reduction by fitting a Cox regression model for each individual, weighted based on their phenotypic similarity to the rest of the trial participants. For the primary outcome of MACE, the median individualised HR (iHR) was 0.63 (IQR 0.53–0.78), with 8800 (94.0%) patients exhibiting an iHR of less than 1, favouring intensive systolic blood pressure treatment (figure 2A). By contrast, for the secondary (net benefit) endpoint of serious adverse events and all-cause mortality, the median iHR was 1.08 (1.01–1.15), with 1988 (21.2%) patient neighbourhoods exhibiting an iHR of less than 1 (figure 2B). In sensitivity analyses, iHRs were poorly correlated with the baseline cardiovascular risk observed in patients in the control group of each phenotypic neighbourhood ($r=-0.04$, 95% CI -0.06 to -0.02), but were moderately to strongly associated with the observed absolute risk reduction ($r=0.65$, 0.64 to 0.66; appendix p 16).

An XGBoost algorithm that predicts a patient's individualised hazard (log HR) based on baseline phenotypic variables, also collected in ACCORD BP, was trained and internally validated in SPRINT. SHAP analysis showed sex, renal function (glomerular filtration rate or creatinine), history of coronary disease requiring revascularisation, history of angina, active smoking, and statin or aspirin use to be key predictors of the individualised MACE benefit of intensive versus standard systolic blood pressure reduction (figure 3).

When we trained an XGBoost tree algorithm on the subset of the eight most important features, there was no evidence of overfitting to SPRINT, with a root mean square error of 0.1930 ($R^2=0.47$) in the 20% of the holdout test set from SPRINT, compared with 0.1864 ($R^2=0.52$) in the training set consisting of 80% of the SPRINT participants. The learning curve for the performance of the XGBoost algorithm in the training and holdout

SPRINT datasets is shown in the appendix (p 17). There was strong correlation between the predictions of the parsimonious (eight variables) and full model ($r=0.97$, 95% CI 0.96–0.97).

Application of the tool in the independent ACCORD BP trial showed that the predicted personalised benefit of intensive systolic blood pressure reduction did not differ between the intensive and standard treatment groups ($p=0.46$), which is consistent with the random allocation of these treatments (appendix p 18). However, individuals with the highest predicted personalised MACE benefit had a lower actual MACE risk when assigned to intensive treatment than to standard treatment, with an adjusted HR for time-to-first MACE of 0.70 (95% CI 0.55–0.90) in individuals with above-median predicted benefit (high responders) versus 1.05 (95% CI 0.84–1.32) for below median predicted benefit (low responders; figure 4, $p_{\text{interaction}}=0.0184$). Nelson-Aalen plots demonstrating the cumulative incidence of MACE in patients with high versus low predicted benefit are shown in figure 4. Among predicted high responders, the hazard curve started to separate 1 year after enrolment in ACCORD BP. These findings were consistent in an analysis where inverse probability censoring weighted estimator accounted for possible dependent censoring (appendix p 19).

Subgroup analysis based on age (<65 years: HR 0.89 [95% CI: 0.71–1.12]; 65 years: 0.85 [0.67–1.09]), sex (male: 0.89 [0.72–1.10]; female: 0.85 [0.65–1.10]), established cardiovascular disease (no: 0.89 [0.70–1.14]; yes: 0.84 [0.67–1.06]) or active smoking (no: 0.85 [0.71–1.02]; yes: 1.01 [0.64–1.60]) did not identify groups with significant benefit from intensive versus standard systolic blood pressure reduction in type 2 diabetes. A screenshot from the browser-accessible version of our decision support tool can be found in the appendix (p 20).

Our positive control simulation studies confirmed the ability of our approach to discover an artificially introduced treatment-effect heterogeneity. The algorithm detected a phenotypic subgroup of patients with significant treatment-effect heterogeneity ($p_{\text{interaction}}<0.0001$); however in a negative control study without any simulated treatment-effect heterogeneity, the algorithm defined no subgroups with significant treatment-effect heterogeneity ($p=0.35$). The effect size detected using our approach was consistent with the effect size introduced in the positive control study (figure 5).

Discussion

Using data from SPRINT, we developed a machine learning-based tool that defines the individualised cardiovascular benefit from intensive versus standard systolic blood pressure reduction based on eight core clinical features: sex, renal function (glomerular filtration rate or creatinine), history of coronary disease requiring revascularisation, history of angina, active smoking, and statin or aspirin use. The tool successfully defined a personalised benefit of intensive blood pressure lowering beyond the SPRINT population in patients with type 2 diabetes in the ACCORD BP trial, although there was no benefit from the approach in the overall study population.⁵ Our algorithm, based on the full breadth and complex inter-relationships of recorded baseline phenotypic information, enabled the extraction of personalised estimates of cardiovascular benefit through iterative analyses of the data

from each participant's unique phenotypic angle. The strategy was robust in simulation experiments with positive and negative control outcomes.

Previous studies have suggested heterogeneous treatment effects in SPRINT defined as an association between an individual's baseline cardiovascular disease risk and the magnitude of absolute risk reduction with intensive versus systolic blood pressure control.^{20–22} To this end, various machine learning approaches—including recursive partition modelling,²³ k-means clustering,²⁴ and X-learners²⁵—have been used to define broad patient groups that experience differential benefit from intensive blood pressure reduction. However, reliance on such broad subgroups and a priori exclusion of variables from the analysis based on presumed lack of clinical significance might restrict the applicability of these approaches.

By contrast, our approach identified complex interactions between a patient's sex, baseline renal function,^{23,24} cardiovascular risk profile (eg, anginal symptoms and previous coronary revascularisation), active smoking, and statin or aspirin use as key determinants of the personalised benefits of targeting a systolic blood pressure goal of less than 120 mm Hg as opposed to less than 140 mm Hg. Our analysis also highlights how differences in the trial-wide outcomes for SPRINT and ACCORD BP might ultimately reflect differences in phenotypes of patients enrolled in the two trials, and therefore supports the cardiovascular benefits of intensive systolic blood pressure reduction in type 2 diabetes,^{26,27} despite the main results of ACCORD BP, which showed no substantial benefits for intensive systolic blood pressure reduction in this population.⁵

We believe that the present work represents a contribution in both methodological and clinical domains. First, our approach treats a trial phenomap as a continuum for evaluating individualised effect estimates. This method expands on our previous work⁹ by yielding robust estimates that incorporate information from all participants in each iteration. Second, in contrast to previous approaches,²³ by modelling the relative risk change, our study provides insights into the phenotypic characteristics that determine the extent to which a patient's baseline cardiovascular risk is modifiable through intensive systolic blood pressure control. Third, our analysis provides a framework for learning from a positive trial to infer patient-level effectiveness in a null trial on hypertension management. Fourth, simulation studies demonstrate the ability of our method to detect meaningful heterogeneous treatment effects in the presence of a ground truth, without overfitting to random noise. Future studies should explore how our approach can complement alternative solutions that have found successful applications across disciplines, such as random forest-based approaches,^{28–30} “U-learners”,³¹ transformed outcome trees,³² or the modified outcome method.³³ Finally, PRECISION, our decision support tool, is aimed at supporting shared decision making between patients and their health-care providers. The PRECISION tool shifts the focus from applying average effects of blood pressure reduction observed in a positive clinical trial to risk reduction for individuals based on their characteristics. Future clinical trials should explore the value of a decision support tool-guided pathway versus traditional pathways in guiding blood pressure management and combating therapeutic inertia.

Our study has a few limitations. First, variation in effect estimates drawn from weighted analyses around each patient's phenotypic location could be prone to random variation.

Future studies should explore the robustness of a dynamic kernel definition, as opposed to a fixed kernel as used in this study. Second, to ensure consistency with the original trials, we chose the original primary outcome as our outcome of interest, with net clinical benefit analyses provided as secondary outcomes that could not generalise across studies due to variable definitions. Third, notable differences in the study design between SPRINT and ACCORD BP—such as the exclusion versus inclusion of patients with diabetes and use of unattended versus attended blood pressure measurements in SPRINT versus ACCORD BP—preclude direct validation of our findings, but support the generalisability of our tool. Fourth, further testing of our tool in diverse patient populations is needed to better understand the biological, clinical, and socioeconomic factors that might underlie such heterogeneous treatment effects. Fifth, our analysis was not compared with other potential methods proposed for the study of heterogeneous treatment effects in trials. However, our approach—which was externally validated in a second independent trial (ACCORD BP) and detected a simulated treatment effect—can be used as a framework to benchmark other methods. Finally, the algorithm should be evaluated prospectively before clinical implementation as a decision support tool.

In this post-hoc analysis of two randomised clinical trials of systolic blood pressure control, we developed a machine learning-guided, evidence-based tool to extract individualised treatment effects based on participant-level data. When applied in two trials of intensive versus standard systolic blood pressure control, our method identified explainable phenotypes associated with heterogeneous treatment effects that generalise to patients without and with type 2 diabetes. More importantly, our broader method provides a personalised approach to the translation of clinical trial findings and promotes the use of shared decision-making through personalised inference.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to the National Heart, Lung, and Blood Institute of the US National Institutes of Health–BioLINCC for providing the clinical trial data, and to all study participants and investigators of SPRINT and ACCORD BP for their contributions to the two studies analysed in this paper. This study was supported by research funding awarded by the Yale School of Medicine to RK. RK also receives support from the National Heart, Lung, and Blood Institute of the US National Institutes of Health (K23HL153775), outside of the submitted work.

Data sharing

Participant-level data of the SPRINT and ACCORD BP trials can be obtained through the National Heart, Lung, and Blood Institute Biologic Specimen and Data Repository Information Coordinating Center. The code for the simulation of the SPRINT trial using the National Heart, Lung, and Blood Institute’s Biologic Specimen and Data Repository Information Coordinating Center data is available through our laboratory’s github page (https://github.com/CarDS-Yale/SPRINT_simulation). Individual requests for access to the code are welcome and can be made through communication with the corresponding author.

The dataset with the simulated outcome will be shared with the BioLINCC team to be posted as an ancillary dataset after publication of the manuscript.

References

1. 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: 795–808. [PubMed: 31492503]
2. Franklin SS, Jacobs MJ, Wong ND, L'Italien GJ, Lapuerta P. Predominance of isolated systolic hypertension among middle-aged and elderly US hypertensives: analysis based on National Health and Nutrition Examination Survey (NHANES) III. *Hypertension* 2001; 37: 869–74. [PubMed: 11244010]
3. Flint AC, Conell C, Ren X, et al. Effect of systolic and diastolic blood pressure on cardiovascular outcomes. *N Engl J Med* 2019; 381: 243–51. [PubMed: 31314968]
4. Wright JT Jr, Williamson JD, Whelton PK, et al. A randomized trial of intensive versus standard blood-pressure control. *N Engl J Med* 2015; 373: 2103–16. [PubMed: 26551272]
5. Cushman WC, Evans GW, Byington RP, et al. Effects of intensive blood-pressure control in type 2 diabetes mellitus. *N Engl J Med* 2010; 362: 1575–85. [PubMed: 20228401]
6. Lewis CE, Fine LJ, Beddhu S, et al. Final report of a trial of intensive versus standard blood-pressure control. *N Engl J Med* 2021; 384: 1921–30. [PubMed: 34010531]
7. Huang C, Dhruva SS, Coppi AC, et al. Systolic blood pressure response in SPRINT (Systolic Blood Pressure Intervention Trial) and ACCORD (Action to Control Cardiovascular Risk in Diabetes): a possible explanation for discordant trial results. *J Am Heart Assoc* 2017; 6: e007509. [PubMed: 29133522]
8. Whelton PK, Carey RM, Aronow WS, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/AphA/ASH/ASPC/NMA/PCNA Guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Hypertension* 2018; 71: 1269–324. [PubMed: 29133354]
9. Oikonomou EK, Van Dijk D, Parise H, et al. A phenomapping-derived tool to personalize the selection of anatomical vs. functional testing in evaluating chest pain (ASSIST). *Eur Heart J* 2021; 42: 2536–48. [PubMed: 33881513]
10. Oikonomou EK, Suchard MA, McGuire DK, Khera R. Phenomapping-derived tool to individualize the effect of canagliflozin on cardiovascular risk in type 2 diabetes. *Diabetes Care* 2022; 45: 965–74. [PubMed: 35120199]
11. Wright MN, Ziegler A. ranger: a fast implementation of random forests for high dimensional data in C++ and R. *J Stat Softw* 2017; 77: 17.
12. Gower JC. A general coefficient of similarity and some of its properties. *Biometrics* 1971; 27: 857–71.
13. McInnes L, Healy J, Melville J. UMAP: uniform manifold approximation and projection for dimension reduction. 2018. arXiv 2018; published online Feb 9. 10.48550/arXiv.1802.03426 (preprint).
14. Schemper M, Wakounig S, Heinze G. The estimation of average hazard ratios by weighted Cox regression. *Stat Med* 2009; 28: 2473–89. [PubMed: 19472308]
15. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. arXiv 2016; published online June 10. 10.48550/arXiv.1603.02754 (preprint).
16. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell* 2020; 2: 56–67. [PubMed: 32607472]
17. Rigdon J, Baiocchi M, Basu S. Preventing false discovery of heterogeneous treatment effect subgroups in randomized trials. *Trials* 2018; 19: 382. [PubMed: 30012181]
18. Willems S, Schat A, van Noorden MS, Fiocco M. Correcting for dependent censoring in routine outcome monitoring data by applying the inverse probability censoring weighted estimator. *Stat Methods Med Res* 2018; 27: 323–35. [PubMed: 26988930]

19. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007; 370: 1453–57. [PubMed: 18064739]
20. Bress AP, Greene T, Derington CG, et al. Patient selection for intensive blood pressure management based on benefit and adverse events. *J Am Coll Cardiol* 2021; 77: 1977–90. [PubMed: 33888247]
21. Patel KK, Arnold SV, Chan PS, et al. Personalizing the intensity of blood pressure control: modeling the heterogeneity of risks and benefits from SPRINT (Systolic Blood Pressure Intervention Trial). *Circ Cardiovasc Qual Outcomes* 2017; 10: e003624. [PubMed: 28373269]
22. Basu S, Sussman JB, Rigdon J, Steimle L, Denton BT, Hayward RA. Benefit and harm of intensive blood pressure treatment: derivation and validation of risk models using data from the SPRINT and ACCORD trials. *PLoS Med* 2017; 14: e1002410. [PubMed: 29040268]
23. Wang S, Khera R, Das SR, et al. Usefulness of a simple algorithm to identify hypertensive patients who benefit from intensive blood pressure lowering. *Am J Cardiol* 2018; 122: 248–54. [PubMed: 29880288]
24. Yang DY, Nie ZQ, Liao LZ, et al. Phenomapping of subgroups in hypertensive patients using unsupervised data-driven cluster analysis: an exploratory study of the SPRINT trial. *Eur J Prev Cardiol* 2019; 26: 1693–706. [PubMed: 31213079]
25. Duan T, Rajpurkar P, Laird D, Ng AY, Basu S. Clinical value of predicting individual treatment effects for intensive blood pressure therapy. *Circ Cardiovasc Qual Outcomes* 2019; 12: e005010. [PubMed: 30857410]
26. Emdin CA, Rahimi K, Neal B, Callender T, Perkovic V, Patel A. Blood pressure lowering in type 2 diabetes: a systematic review and meta-analysis. *JAMA* 2015; 313: 603–15. [PubMed: 25668264]
27. Turnbull F, Neal B, Algert C, et al. Effects of different blood pressure-lowering regimens on major cardiovascular events in individuals with and without diabetes mellitus: results of prospectively designed overviews of randomized trials. *Arch Intern Med* 2005; 165: 1410–19. [PubMed: 15983291]
28. Athey S, Wager S. Estimating treatment effects with causal forests: an application. arXiv 2019; published online Feb 20. 10.48550/arXiv.1902.07409 (preprint).
29. Yao L, Chu Z, Li S, Li Y, Gao J, Zhang A. A survey on causal inference. arXiv 2020; published online Feb 5. 10.48550/arXiv.2002.02770 (preprint).
30. Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests. arXiv 2015; published online Oct 14. 10.48550/arXiv.1510.04342 (preprint).
31. Nie X, Wager S. Quasi-Oracle estimation of heterogeneous treatment effects. arXiv 2017; published online Dec 13. 10.48550/arXiv.1712.04912 (preprint).
32. Athey S, Imbens G. Recursive partitioning for heterogeneous causal effects. *Proc Natl Acad Sci USA* 2016; 113: 7353–60. [PubMed: 27382149]
33. Tian L, Alizadeh AA, Gentles AJ, Tibshirani R. A simple method for estimating interactions between a treatment and a large number of covariates. *J Am Stat Assoc* 2014; 109: 1517–32. [PubMed: 25729117]

Research in context

Evidence before this study

Hypertension is a key modifiable cardiovascular risk factor. Although the effects of blood pressure reduction in hypertension are well established, the optimal treatment goals for individual patients remain under debate. We searched PubMed on July 19, 2021, for studies published in English and relating to decision support tools to guide the personalisation of systolic blood pressure treatment targets. A second search was performed on Sept 29, 2022, with the same search criteria, to update our literature review. We used the search terms “hypertension”, “intensive blood pressure control”, “systolic blood pressure”, “heterogeneity”, and “machine learning”. Our search returned 5278 papers, 29 of which were relevant to the topic. Two large randomised clinical trials have evaluated the effect of intensive systolic blood pressure lowering for cardiovascular risk reduction. The SPRINT trial found that targeting a systolic blood pressure of less than 120 mm Hg (intensive treatment) as opposed to less than 140 mm Hg (standard treatment) reduces the incidence of major adverse cardiovascular events among patients with hypertension but without diabetes (types 1 and 2). By contrast, the ACCORD BP trial did not show a cardiovascular benefit for the intensive versus standard treatment strategy in type 2 diabetes. Our search revealed several studies and approaches for detecting and describing heterogeneity in the effects of intensive systolic blood pressure treatment. To date, studies have used a mixture of approaches, ranging from regression modelling with interaction effects to machine learning algorithms. Although many studies have suggested that the absolute risk reduction with intensive systolic blood pressure control is proportional to the baseline cardiovascular disease risk, others have demonstrated that more advanced machine learning algorithms might improve the detection of individualised treatment effects. Nevertheless, most previous studies restricted their analysis to modelling of absolute risk reduction, relied on a smaller number of recorded covariates, defined discrete numbers of phenotypic clusters, or did not demonstrate generalisability of their approach to independent patient populations.

Added value of this study

In this post-hoc analysis of SPRINT and ACCORD BP, we used machine learning to construct computational clinical trial phenomaps and investigate patient-level heterogeneity in treatment effectiveness of intensive blood pressure control. We leveraged this heterogeneity to develop an evidence-based tool to personalise the consideration for pursuing intensive versus standard systolic blood pressure treatment goals among patients without and with type 2 diabetes with high cardiovascular risk. This individualised approach to evidence synthesis could represent a novel strategy to maximise the benefits of intensive blood pressure control and might be a valuable adjunct to inform shared decision making in clinical practice. The approach also highlights a potential strategy for personalised inference from randomised clinical trials.

Implications of all the available evidence

Here, we present an objective decision support tool derived from two of the largest randomised clinical trials, to date, to have assessed the cardiovascular benefit of intensive

versus standard systolic blood pressure reduction in hypertension. This tool might facilitate a more standardised, yet personalised, approach in the selection of optimal treatment goals for individual patients. Our analyses demonstrate that there is substantial heterogeneity in the individual benefit derived from pursuing intensive blood pressure control, and that phenotypic features that define benefit from intensive systolic blood pressure reduction might generalise to patients without and with type 2 diabetes.

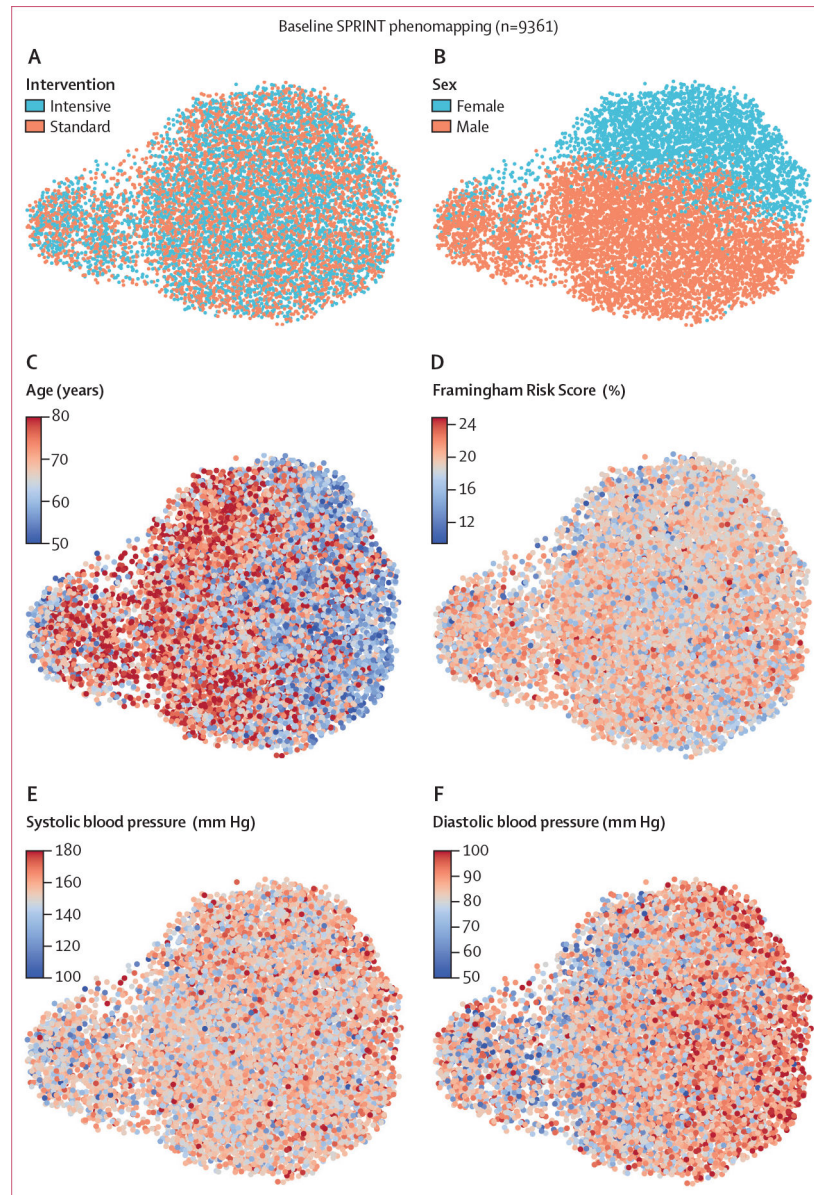


Figure 1: Manifold representations of the phenotypic architecture of SPRINT
 Patients are embedded in the phenotypic space based on dissimilarity metrics (Gower's distance) derived from 59 pre-randomisation variables; thus phenotypically similar individuals tend to be topologically closer. Each dot represents a study participant, with colouring based on treatment group. Since the dimensionality reduction is non-linear, axes have been omitted and only the comparisons between distances are meaningful. SPRINT=Systolic Blood Pressure Intervention Trial.

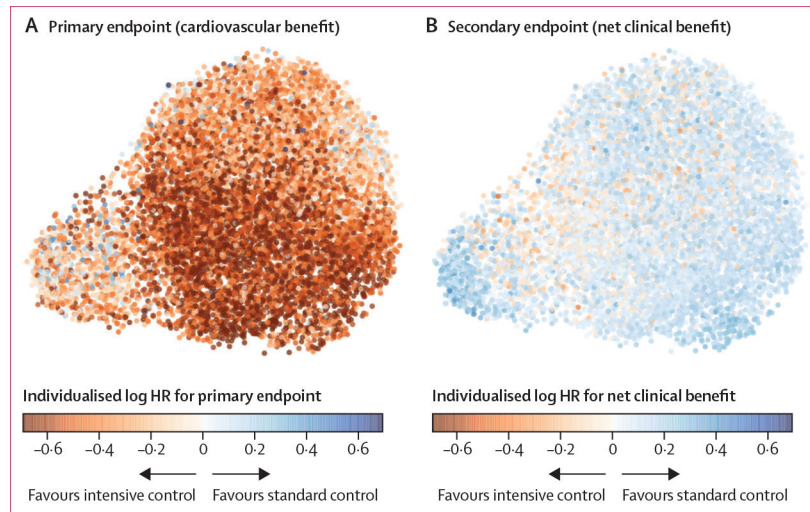


Figure 2: Cardiovascular benefit phenomaps of intensive blood pressure reduction in SPRINT
phenomap representation of the individualised HRs with intensive versus standard blood
pressure control for the primary (A) and secondary (B) outcomes
 HR=hazard ratio. SPRINT=Systolic Blood Pressure Intervention Trial.

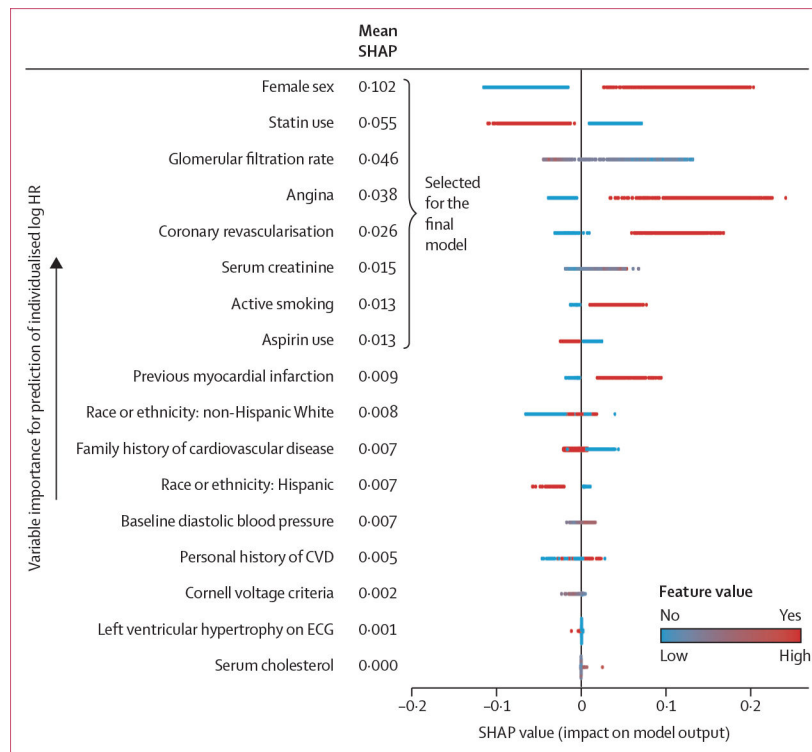


Figure 3: SHAP analysis showing feature importance for prediction of individualised cardiovascular benefit from intensive systolic blood pressure reduction

The y-axis represents the features included in the model development (in descending order of importance) and the x-axis indicates the change in prediction. The gradient colour denotes the original value for that variable (eg, for categorical variables such as sex it only takes two colours, whereas for continuous variables it contains the whole spectrum), with each point representing an individual participant from SPRINT. More negative SHAP values indicate a higher major adverse cardiovascular event benefit with an intensive versus standard treatment strategy. The eight most important variables (with importance of 0.01 or higher) were selected to train a parsimonious clinical model. CVD=cardiovascular disease. HR=hazard ratio. SHAP=Shapley additive explanations. SPRINT=Systolic Blood Pressure Intervention Trial.

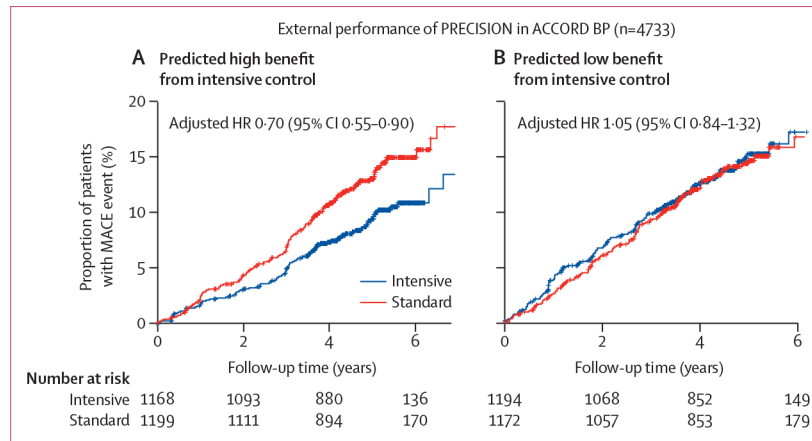


Figure 4: External performance of the PRECISION decision support tool in ACCORD BP, a trial of intensive systolic blood pressure reduction among patients with type 2 diabetes Nelson-Aalen plots demonstrating the cumulative incidence of MACE in patients with high (above the median; A) versus low (below the median; B) predicted benefit. Models were adjusted for age and sex. ACCORD BP=Action to Control Cardiovascular Risk in Diabetes Blood Pressure trial. HR=hazard ratio. MACE=major adverse cardiovascular event.

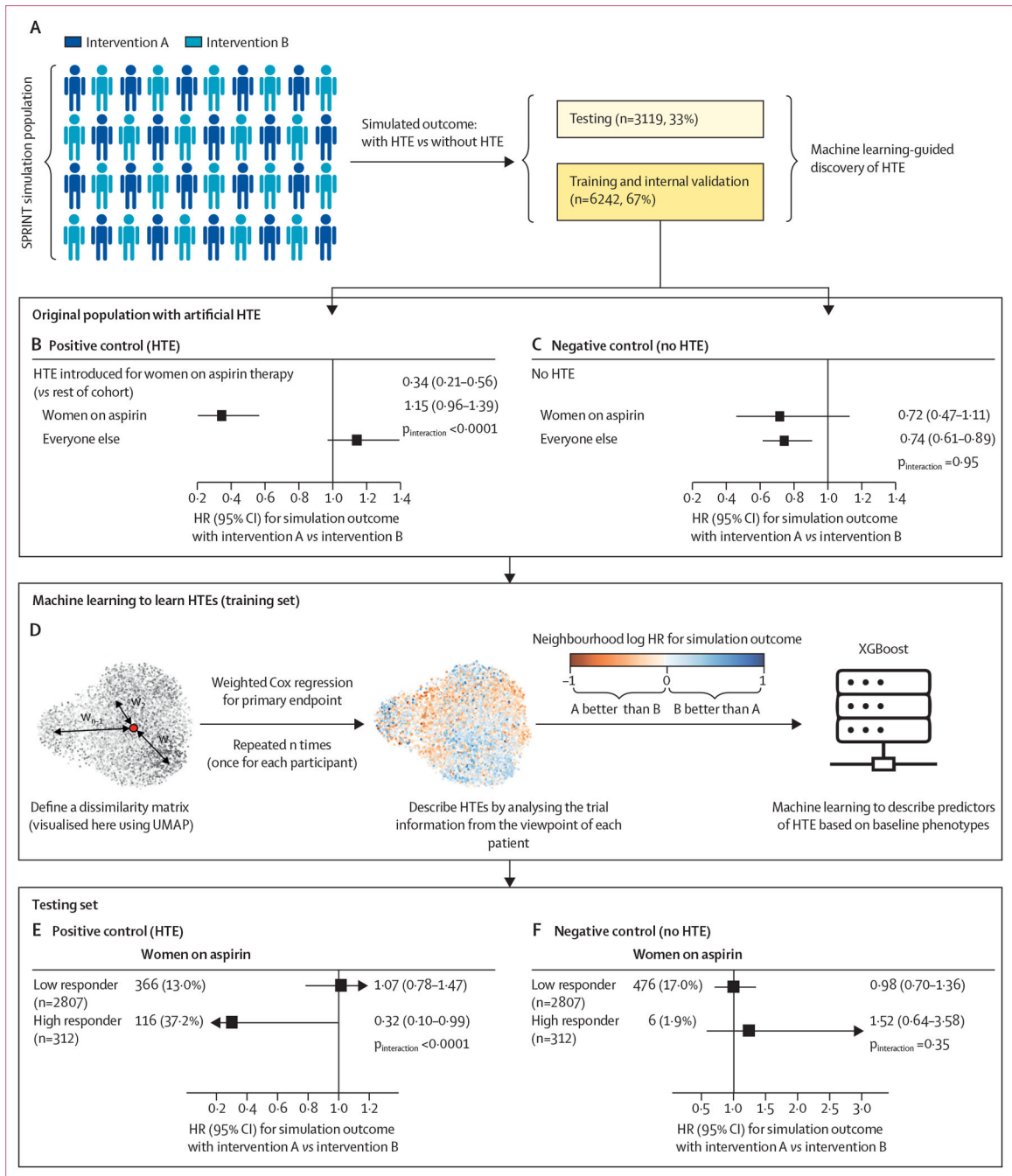


Figure 5: Simulation studies for sensitivity analysis

(A) The SPRINT population was randomly split into a training and internal validation set (n=6242, 67%) and a testing set (n=3119, 33%). In the positive control simulation, with HTE (B), an average treatment effect of 6% was introduced among women receiving aspirin. In the negative control simulation (no HTE; C), an average treatment effect of 1% was introduced across the population. After defining a dissimilarity matrix and phenomap for the training population and extracting individualised hazard estimates for each participant, an XGBoost model was trained to detect baseline phenotypes associated

with HTEs (D). In the held-out test set, the algorithm successfully identified patient groups with HTE in the positive (E) but not the negative (F) control simulations. HR=hazard ratio. HTE=heterogeneous treatment effect. SPRINT=Systolic Blood Pressure Intervention Trial. UMAP=uniform manifold approximation. XGBoost=extreme gradient-boosting algorithm.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript