Published in final edited form as:

Ann Intern Med. 2017 March 07; 166(5): 354–360. doi:10.7326/M16-1756.

Detecting Heterogeneous Treatment Effects to Guide Personalized Blood Pressure Treatment:

A Modeling Study of Randomized Clinical Trials

Sanjay Basu, MD, PhD, Jeremy B. Sussman, MD, MS, and Rod A. Hayward, MD

Stanford University, Stanford, California; Harvard Medical School, Boston, Massachusetts; and University of Michigan, Ann Arbor, Michigan

Abstract

Background—Two recent randomized trials produced discordant results when testing the benefits and harms of treatment to reduce blood pressure (BP) in patients with cardiovascular disease (CVD).

Objective—To perform a theoretical modeling study to identify whether large, clinically important differences in benefit and harm among patients (heterogeneous treatment effects [HTEs]) can be hidden in, and explain discordant results between, treat-to-target BP trials.

Design—Microsimulation.

Data Sources—Results of 2 trials comparing standard (systolic BP target <140 mm Hg) with intensive (systolic BP target <120 mm Hg) BP treatment and data from the National Health and Nutrition Examination Survey (2013 to 2014).

Target Population—U.S. adults.

Time Horizon—5 years.

Perspective—Societal.

Intervention—BP treatment.

Requests for Single Reprints: Sanjay Basu, MD, PhD, Center for Population Health Sciences, Stanford University School of Medicine, 1070 Arastradero Road 282, MC 5560, Palo Alto, CA 94304; basus@stanford.edu.

Current Author Addresses: Dr. Basu: Center for Population Health Sciences, Stanford University School of Medicine, 1070 Arastradero Road 282, MC 5560, Palo Alto, CA 94304.

Dr. Sussman: University of Michigan, 2800 Plymouth Road, Building 16, Room 335E, Ann Arbor, MI 48109.

Dr. Hayward: University of Michigan, 2800 Plymouth Road, Building 10, Room G016-4A, Ann Arbor, MI 48109.

Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the U.S. Department of Veterans Affairs.

Disclosures: Authors have disclosed no conflicts of interest. Forms can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M16-1756.

Reproducible Research Statement: *Study protocol:* Not applicable. *Statistical code and data set:* Available at sdr.stanford.edu.

Author Contributions: Conception and design: J.B. Sussman, R.A. Hayward.

Analysis and interpretation of the data: S. Basu, R.A. Hayward.

Drafting of the article: S. Basu.

Critical revision of the article for important intellectual content: J.B. Sussman, R.A. Hayward.

Final approval of the article: S. Basu, J.B. Sussman, R.A. Hayward.

Statistical expertise: R.A. Hayward.

Administrative, technical, or logistic support: R.A. Hayward.

Outcome Measures—CVD events and mortality.

Results of Base-Case Analysis—Clinically important HTEs could explain differences in outcomes between 2 trials of intensive BP treatment, particularly diminishing benefit with each additional BP agent (for example, adding a second agent reduces CVD risk [hazard ratio, 0.61], but adding a fourth agent to a third has no benefit) and increasing harm at low diastolic BP.

Results of Sensitivity Analysis—Conventional treat-to-target trial designs had poor (<5%) statistical power to detect the HTEs, despite large samples ($n > 20\,000$), and produced biased effect estimates. In contrast, a trial with sequential randomization to more intensive therapy achieved greater than 80% power and unbiased HTE estimates, despite small samples ($n = 3500$).

Limitations—The HTEs as a function of the number of BP agents only were explored. Simulated aggregate data from the trials were used as model inputs because individual-participant data were not available.

Conclusion—Clinically important heterogeneity in intensive BP treatment effects remains undetectable in conventional trial designs but can be detected in sequential randomization trial designs.

Primary Funding Source—National Institutes of Health and U.S. Department of Veterans Affairs.

High blood pressure (BP) remains a leading risk factor for cardiovascular disease (CVD) in the United States (1). Recently, 2 large randomized trials investigated the effect of treating patients with high BP and high CVD risk who received intensive therapy (systolic BP [SBP] target <120 mm Hg) or standard therapy (SBP target <140 mm Hg). The results of the 2 trials differed in their CVD mortality outcomes. In the ACCORD (Action to Control Cardiovascular Risk in Diabetes) study (2), published in 2010 ($n = 4733$), patients randomly assigned to intensive therapy, compared with standard therapy, did not have reductions in the primary composite outcome of CVD events (hazard ratio [HR], 0.88 [95% CI, 0.73 to 1.06]) or CVD mortality (HR, 1.06 [CI, 0.74 to 1.52]). In SPRINT (Systolic Blood Pressure Intervention Trial) (3), published in 2015 ($n = 9361$), patients randomly assigned to intensive therapy had both clinically and statistically significant reductions in CVD events (HR, 0.75 [CI, 0.64 to 0.89]) and CVD mortality (HR, 0.57 [CI, 0.38 to 0.85]). Although the relative risk reduction in CVD events observed in the trials (12% in ACCORD vs. 25% in SPRINT) did not differ significantly, differences in the relative risk reduction in CVD mortality were statistically significant (6% increase in ACCORD vs. 43% decrease in SPRINT; $P < 0.05$) (Supplement Table 1, available at Annals.org).

Experts interpreting the results of these 2 trials have proposed several theories for the differences. Although the results may differ by chance alone, this seems unlikely given the large sample sizes and large differences in effects observed between the trials. Another possible explanation is that persons with diabetes (enrolled in ACCORD) may not benefit from intensive BP control, but those without diabetes (enrolled in SPRINT) do. However, prior studies discount this possibility (4). For example, the HOT (Hypertension Optimal Treatment) trial ($n = 18\,790$) (5) found benefits from intensive therapy only among participants with diabetes. Alternative theories attribute the differences in trial outcomes to

differences in other sample characteristics (the average age of participants was greater in SPRINT, which included those with chronic kidney disease—a group that was excluded from ACCORD), outcome definitions (heart failure was included in the overall CVD composite outcome in SPRINT but not ACCORD), or BP measurement technique (SPRINT protocols used a different method of measurement). However, subgroup analyses in SPRINT revealed similar benefits of intensive therapy across ages and kidney function levels that persisted when heart failure was excluded from the outcome definition (3). Because these analyses are typically underpowered and limited to studying 1 clinical feature at a time (rather than combinations of such features as age, biomarker levels, and CVD risk levels) (6 – 12), clinicians are often left with the dilemma of how best to interpret these conflicting results.

The inability to concretely identify why ACCORD and SPRINT results differ is emblematic of the finding that conventional treat-to-target trials have difficulty providing reliable evidence to decipher heterogeneous treatment effects (HTEs) (13); in other words, to identify whether and why some patients have greater or fewer benefits or risks from a therapeutic decision. Uncovering HTEs is critical to clinicians because benefit and risk assessments are required to ensure that individual patients receive the benefits observed in SPRINT while avoiding the potential harms observed in ACCORD.

Two specific causes of HTEs, the number of BP agents provided and the diastolic BP (DBP) achieved, might explain the differences between the trials' outcomes. Treatment in these studies differed markedly in how many BP agents the intensive therapy group received (41% received 4 agents in ACCORD vs. 24% in SPRINT) (Table) and DBP achieved (64 mm Hg in ACCORD vs. 69 mm Hg in SPRINT) (2, 3). Very low DBP has been associated with increased mortality, independent of SBP, in subgroup analyses of both trials and cohort studies (14); analogously, decreasing benefit with increasing medication intensity has been observed because of diminishing effectiveness and increasing polypharmacy (15). However, we cannot know whether these effects are real, because subgroup analysis of trials neutralize the benefits of randomization, which leads to potentially biased results (16).

We hypothesize that to decipher such HTEs, a SMART (sequential, multiple assignment, randomized trial) may be useful (17). A SMART design (Figure 1) randomly assigns participants to either be intensified or not be intensified to each additional agent. Hence, the design allows us to examine effects of treatment intensification along the pathway of achieving a target (for example, SBP <120 mm Hg) rather than only compare aggregate outcomes among intensive therapy and control groups. The design is thus intended to overcome a major problem with conventional designs: Participants in each group receive varied treatments in pursuit of the assigned treatment goals (for example, to achieve the target a person in the intensive therapy group could require only 1 medication and someone else could require 4). Heterogeneous treatments in a study group can make it difficult to identify whether some strategies conferred benefit among some participants and harm among others (18).

Using microsimulation models, we performed a thought experiment to determine whether a priori-hypothesized HTEs with strong prior evidence could be hidden in, and explain the

conflicting results between, conventional treat-to-target trials. We analyzed whether a conventional treat-to-target trial design, as used in ACCORD and SPRINT, could reliably and accurately detect such HTEs if they exist and, alternatively, whether a SMART design could better identify these potentially hidden HTEs.

Methods

Different amounts and sources of heterogeneity can and probably do exist within trials, independent of the differences or similarities of their overall treatment effects (11). Our objective was to investigate whether large, clinically important HTEs could be hidden in the trials and explain their discordance in overall treatment effects, not what all possible HTEs might be. We provide an overview of the methods, with complete details in the Supplement (available at Annals.org).

Microsimulation Model of ACCORD and SPRINT

To examine whether HTEs could be hidden in the ACCORD and SPRINT trials and could explain the discordant results, we developed a microsimulation model to explain the results of both trials. The model was designed to test the following 2 primary HTEs: treatment to very low DBP could pose harm, offsetting the benefit of treating to lower SBP; and use of many BP agents could pose harm, also offsetting this benefit of treating to lower SBP. The model simulated each participant in each group of each trial, their reported clinical features and BP agents received, and subsequent CVD events and deaths (Table). Because individual-participant trial data were not available for this analysis, we simulated the patient characteristics (pre- and posttreatment SBP and DBP, Framingham risk score, and number of agents used) to match summary statistics from each trial publication. We also performed repeated simulations across the range of input values reported to capture the range of uncertainty. The CVD events were defined as coronary death, myocardial infarction, coronary insufficiency, angina, ischemic stroke, hemorrhagic stroke, transient ischemic attack, peripheral artery disease, or heart failure.

We created 3 simple models incorporating only a few key variables supported by prior theory or empirical evidence. Specifically, our models estimated the HR of CVD events and CVD deaths between the 2 groups of each trial as a function of 5 variables: pretreatment and posttreatment SBP, posttreatment DBP, number of BP agents used, and 10-year pretreatment Framingham CVD risk score (19). In model 1, we considered all 5 variables. Next, we calibrated models that omitted the number of BP agents used (model 2) and another that omitted DBP after treatment (model 3). We calibrated each model to find the variable coefficients that allowed it to simultaneously match the HRs for CVD events and deaths for both trial populations. Using the calibrated models, we estimated the degree to which individual participants receiving intensive therapy could differ in their expected benefits and risks for intensification and still fit both sets of trial results.

SMART Design Simulation

We then simulated a SMART trial with sequential randomization to increasingly intensive BP therapy (Figure 1) to estimate potential benefits and risks at each stage of intensification.

This analysis used aggregate trial participant data and models calibrated to ACCORD and SPRINT and specified the BP reductions for each medication based on a prior meta-analysis (20). Inclusion criteria were age 50 to 75 years, 10-year Framingham CVD risk score of 10% or greater, SBP of 135 to 150 mm Hg and DBP of 65 to 100 mm Hg when receiving 1 or 2 BP agents, and no history of congestive heart failure. The correlations among age, risk score, SBP, and DBP were taken from the National Health and Nutrition Examination Survey (2013 to 2014) to account for the correlations between risk factors (that is, increased SBP with age) (21).

Three randomizations were simulated (Figure 1): a subset of participants receiving 1 BP agent were randomly assigned to either continue their regimen or add a second agent; those receiving 2 agents who did not achieve SBP less than 120 mm Hg were randomly assigned to either continue their regimen or add a third agent; and those receiving 3 BP agents who did not achieve an SBP less than 120 mm Hg were randomly assigned to either continue their regimen or add a fourth agent. Hence, the SMART design still supports treatment to the SBP target of 120 mm Hg, but it randomly assigned participants to increase or continue their number of BP agents in pursuit of the target.

Eligibility for randomization to the next more intensive phase of therapy was based on the average of 2 simulated measurements of BP (with a coefficient of variation of 0.09 to capture measurement error and biological variation across clinical recordings) (22, 23). We adopted an intention-to-treat approach to estimate HRs of treatment effect. Our primary outcome metric was the HR of CVD events from each stage of intensified treatment (Supplement). The secondary outcome metric was CVD mortality.

Power Calculations for Treat-to-Target and SMART Designs

To examine how large a trial would need to be to achieve greater than 80% power to identify clinically important HTEs, we simulated conventional treat-to-target and SMART trials with sample sizes varying from 2000 to 20 000 participants. Concordant with current recommendations for detecting HTEs (11, 24), power was defined as the percentage of 10 000 repeated simulations at a given sample size that produced a significant logistic regression coefficient (2-tailed $P < 0.05$ level) for the interaction term of the treatment group (0 if standard or 1 if intensive) and 1 of the following: number of BP agents before intensification (1, 2, or 3) or DBP less than 70 mm Hg (0 if >70 mm Hg) before intensification. For each of 10 000 trials simulated at each sample size, we sampled with replacement from the probability distributions of all input parameters (Supplement) to estimate the degree of random variation and uncertainty in the outcome rates. We then compared the estimated HTE size and power of the conventional treat-to-target trials with the SMART trials at each sample size. All modeling was performed in R, version 3.1.2 (R Foundation for Statistical Computing).

The study was deemed exempt from ethics panel review by the Stanford University Institutional Review Board (e-protocol ID 34359).

Role of the Funding Source

The funders had no role in conceiving, designing, or performing the study; drafting or editing the manuscript; or interpreting, presenting, or deciding to submit the results for publication.

Results

Microsimulation Model of ACCORD and SPRINT

Using the trial characteristics of each study group (Table), our microsimulation models showed that clinically important HTEs by DBP and number of agents could explain the differences in CVD events and CVD deaths between ACCORD and SPRINT. As shown in Figure 2, the fitted HTE equations predicted that the benefits of intensified therapy decreased when posttreatment DBP was less than 70 mm Hg or more BP agents were added. (See the Supplement for full equations, which also included variables for pretreatment and posttreatment SBP and pretreatment CVD risk.) For example, a person treated to 60 mm Hg would be expected to have an HR of CVD events that is 0.08 higher than someone treated only to 65 mm Hg. Further, a person requiring 4 BP agents would be expected to have an HR of CVD events that is 0.20 higher than someone achieving the same SBP target with 3 BP agents. We illustrated the DBP of 70 mm Hg as a prespecified cut point to match prior BP-lowering trials (14). The HTE by pretreatment CVD risk was small because of limited variation in CVD risk among the 2 trials in which all participants were at high risk. Supplement Figure 1 shows that our models closely approximated both the mean and CIs around the HRs of CVD events and CVD deaths in both trials. Alternative models that examined HTE by low DBP (model 2) or number of BP agents (model 3), but not both, could similarly explain the differences between ACCORD and SPRINT results.

Supplement Table 2 shows how clinically important the amount of HTE reported in Figure 2 would be, if present. For example, treatment intensification would produce great benefit in a patient with high CVD risk and a BP of 135/75 mm Hg who received no BP agents before intensification (1 CVD event prevented for every 9 to 19 patients treated for 5 years with intensive SBP control). But treatment intensification would cause great harm in a patient with moderate CVD risk and refractory systolic hypertension (1 CVD death caused for every 25 to 135 patients treated for 5 years with intensive SBP control).

Statistical Power to Detect HTE of a Conventional Clinical Trial Versus SMART Design

These results only demonstrate that clinically important HTE could explain the differences in ACCORD and SPRINT, but they cannot address whether true, clinically important HTEs could reliably be found in these trials. Therefore, we next examined whether the large HTEs in Figure 2 would probably be detected by subgroup analyses of conventional treat-to-target trial designs, such as those used for ACCORD and SPRINT. As shown in Figure 3, the conventional treat-to-target trial designs would have less than 5% power to detect the HTEs, even with sample sizes of 20 000 participants (2.9% power for DBP and 3.9% power for number of BP agents). In simpler models in which we limited major HTE to only low DBP (Supplement Figure 3, *A*) or only the number of BP agents (Supplement Figure 3, *B*), the power did not substantially improve.

In contrast, we found that a SMART trial design would have greater than 80% power to detect HTE by the number of BP agents with a sample size as small as 3500 participants (Figure 3). This type of trial would need to have a substantially larger sample size of 16 200 participants to have greater than 80% power to detect the theorized HTE by DBP (<70 mm Hg). If a single HTE was dominant (Supplement Figure 3, *A* and *B*), the required sample sizes would decrease modestly (approximately 14 700 to detect HTE in DBP and 3400 to detect HTE in the number of agents).

Further analysis identified the main reason that conventional treat-to-target trials have less power to detect HTEs. Each study group has a wide range of starting values for DBP and BP agents. This made it difficult for the conventional design to detect which factors in each group were contributing to benefits or harms (that is, the participants in each group were very different, which reduced the ability to have sufficient power to identify the HTE) (Supplement Figure 4). In addition, the proportion of participants who were randomly assigned to a similar number of additional BP agents but differed in their prerandomization DBP or prerandomization number of BP agents was severely imbalanced, which reduced the effective sample size to detect the HTEs in the treat-to-target design (Supplement Figure 5). Specifically, the simulations showed that no participants in the standard therapy group received 3 or more additional agents and no participants in the intensive therapy group received 0 additional agents or only 1 additional agent. Thus, the only contrast involving participants in both groups was among those who received 2 additional agents (Supplement Figure 5). In contrast, the SMART trial randomly assigned similar participants to more or less intensive therapy. This allowed for better identification of which factors contributed to benefits or harms and a more balanced sample size during comparison of participants who were randomly assigned to a similar number of additional BP agents but differed with respect to prerandomization DBP or number of BP agents before intensification. Of note, the conventional treat-to-target trial design produced estimates biased toward the null for HTE sizes. In contrast, the SMART trial design produced unbiased point estimates of the HTE sizes (Supplement Figure 6).

Discussion

Traditional treat-to-target trial designs have great intuitive appeal. Participants are randomly assigned to treatment targets, such as a specific SBP goal, and agents are advanced in pursuit of that goal. If the trial is successful and subgroup analyses fail to show HTEs, then it is generally assumed that the results can be safely generalized to all persons who met the trial's inclusion criteria (3). Our findings raise doubts about these assumptions. We found that large and clinically important HTEs can be hidden in conventional treat-to-target designs. Trial designs that use sequential randomization are much better at detecting such HTEs and potentially provide better evidence to guide personalized treatment. Conflicting results from treat-to-target trials are not limited to BP studies. They also are found in diverse fields in which treatment to biomarker levels is used to guide such treatment as that of type 2 diabetes to achieve a specific hemoglobin A_{1c} target (25).

The primary issue with the conventional treat-to-target design is the heterogeneity of the treatment received by patients in the same study group and its potential influence on

outcomes. For example, in SPRINT's intensive therapy group, some patients received 2 agents whereas others ultimately required at least 4 agents. Similarly, in ACCORD's intensive therapy group, some patients had DBP decreased to less than 55 mm Hg and others had a final DBP greater than 75 mm Hg. A treat-to-target trial design cannot reliably examine HTEs. Because these heterogeneous treatment advances occur after randomization, we cannot match intensive treatment patients who needed fewer or more agents to achieve the SBP goal to the appropriate control patients.

Our analysis focused on 2 of the largest, most rigorously designed trials of SBP treatment intensification that had similar designs but produced different results. However, our general findings are relevant to all randomized, controlled trials that involve treatment intensification protocols. Subgroup analyses, including those in individual-level meta-analyses, have known limitations in helping clinicians translate their results to individual patients (6–12). Our results further demonstrate that treat-to-target randomized, controlled trials have inherent, yet poorly appreciated, additional deficiencies that can hide HTEs that occur in the treatment pathway. Of note, we observed that a trial with reasonable sample size and sequential randomization can detect hypothesized HTEs by deciphering risk and benefit along the pathway of treatment intensification. In addition, a treat-to-target trial would not be expected to detect large and clinically important HTEs. It would be expected to produce biased HTE size estimates if HTEs were detected at all.

As with any study relying on modeling, our study is limited by its assumptions. We explored only HTEs by the number of agents, not by drug classes, and we did not include the possibility of independent effects of diabetes on CVD risk or the importance of different choices in SBP measurement or CVD outcome definitions. Greater access to individual-level trial data may help test some of HTEs theorized in our study, which are based on summary statistics alone. However, we found that HTE analysis of individual-level trial data from a treat-to-target trial would likely be biased. Of note, treat-to-target designs are inherently limited for assessing whether HTEs are hidden in such trials, which makes explaining why trials yielded different results potentially difficult.

Although all HTE examinations require a priori specification (12, 23), sequential randomization is uniquely capable of allowing the examination of such issues. The SMART design may be helpful in analyzing HTEs and identifying which patients may benefit or face risks as a target is pursued. Our trial design can detect HTEs in characteristics that can be randomized (such as the number of BP agents) or are directly affected by the characteristic being randomized (such as DBP), but not necessarily all HTEs. However, the SMART design would add complexity to the conduct of clinical trials and would be less straightforward to analyze.

In conclusion, large and clinically important HTEs, such as benefit decreasing as more BP agents are added or harm occurring from low DBP, could be hidden in the SPRINT and ACCORD trials and could explain their discordant results. Trials using a treat-to-target design would have very poor statistical power to detect such HTEs and thus would produce biased results. In contrast, a large clinical trial using a sequential randomization design would have good to excellent statistical power for detecting and correctly estimating HTEs

and thus would provide better clinical evidence for optimal, personalized prescribing of BP agents.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Grant Support: By the National Institutes of Health (awards DP2MD010478, U54MD010724, K08HL121056, and P60DK 20572) and the U.S. Department of Veterans Affairs Health Services Research & Development Service (award IIR02225).

References

1. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, et al. American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics—2015 update: a report from the American Heart Association. *Circulation*. 2015; 131:e29–322. DOI: 10.1161/CIR.000000000000152 [PubMed: 25520374]
2. Cushman WC, Evans GW, Byington RP, Goff DC Jr, Grimm RH Jr, Cutler JA, et al. ACCORD Study Group. Effects of intensive blood-pressure control in type 2 diabetes mellitus. *N Engl J Med*. 2010; 362:1575–85. DOI: 10.1056/NEJMoa1001286 [PubMed: 20228401]
3. Wright JT Jr, Williamson JD, Whelton PK, Snyder JK, Sink KM, Rocco MV, et al. SPRINT Research Group. A randomized trial of intensive versus standard blood-pressure control. *N Engl J Med*. 2015; 373:2103–16. DOI: 10.1056/NEJMoa1511939 [PubMed: 26551272]
4. Turnbull F, Neal B, Algert C, Chalmers J, Chapman N, Cutler J, et al. Blood Pressure Lowering Treatment Trialists' Collaboration. 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; 165:1410–9. [PubMed: 15983291]
5. Hansson L, Zanchetti A, Carruthers SG, Dahlöf B, Elmfeldt D, Julius S, et al. Effects of intensive blood-pressure lowering and low-dose aspirin in patients with hypertension: principal results of the Hypertension Optimal Treatment (HOT) randomised trial. HOT Study Group. *Lancet*. 1998; 351:1755–62. [PubMed: 9635947]
6. Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey Smith G. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. *Health Technol Assess*. 2001; 5:1–56.
7. Albert JM, Gadbury GL, Mascha EJ. Assessing treatment effect heterogeneity in clinical trials with blocked binary outcomes. *Biom J*. 2005; 47:662–73. [PubMed: 16385907]
8. Rothwell PM, Warlow CP. Prediction of benefit from carotid endarterectomy in individual patients: a risk-modelling study. European Carotid Surgery Trialists' Collaborative Group. *Lancet*. 1999; 353:2105–10. [PubMed: 10382694]
9. Hayward RA, Kent DM, Vijan S, Hofer TP. Multivariable risk prediction can greatly enhance the statistical power of clinical trial subgroup analysis. *BMC Med Res Methodol*. 2006; 6:18. [PubMed: 16613605]
10. Brookes ST, Whitley E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epidemiol*. 2004; 57:229–36. [PubMed: 15066682]
11. Burke JF, Hayward RA, Nelson JP, Kent DM. Using internally developed risk models to assess heterogeneity in treatment effects in clinical trials. *Circ Cardiovasc Qual Outcomes*. 2014; 7:163–9. DOI: 10.1161/CIRCOUTCOMES.113.000497 [PubMed: 24425710]
12. Burke JF, Sussman JB, Kent DM, Hayward RA. Three simple rules to ensure reasonably credible subgroup analyses. *BMJ*. 2015; 351:h5651.doi: 10.1136/bmj.h5651 [PubMed: 26537915]

13. Hayward RA, Hofer TP, Vijan S. Narrative review: lack of evidence for recommended low-density lipoprotein treatment targets: a solvable problem. *Ann Intern Med.* 2006; 145:520–30. [PubMed: 17015870]
14. Banach M, Aronow WS. Blood pressure j-curve: current concepts. *Curr Hypertens Rep.* 2012; 14:556–66. DOI: 10.1007/s11906-012-0314-3 [PubMed: 23054894]
15. Timbie JW, Hayward RA, Vijan S. Diminishing efficacy of combination therapy, response-heterogeneity, and treatment intolerance limit the attainability of tight risk factor control in patients with diabetes. *Health Serv Res.* 2010; 45:437–56. DOI: 10.1111/j.1475-6773.2009.01075.x [PubMed: 20070387]
16. VanderWeele TJ, Knol MJ. Interpretation of subgroup analyses in randomized trials: heterogeneity versus secondary interventions. *Ann Intern Med.* 2011; 154:680–3. DOI: 10.7326/0003-4819-154-10-201105170-00008 [PubMed: 21576536]
17. Lei H, Nahum-Shani I, Lynch K, Oslin D, Murphy SA. A “SMART” design for building individualized treatment sequences. *Annu Rev Clin Psychol.* 2012; 8:21–48. DOI: 10.1146/annurevclinpsy-032511-143152 [PubMed: 2224838]
18. Chakraborty B, Murphy SA. Dynamic treatment regimes. *Annu Rev Stat Appl.* 2014; 1:447–64. DOI: 10.1146/annurev-statistics-022513-115553 [PubMed: 25401119]
19. D’Agostino RB Sr, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation.* 2008; 117:743–53. DOI: 10.1161/CIRCULATIONAHA.107.699579 [PubMed: 18212285]
20. Law MR, Wald NJ, Morris JK, Jordan RE. Value of low dose combination treatment with blood pressure lowering drugs: analysis of 354 randomised trials. *BMJ.* 2003; 326:1427. [PubMed: 12829555]
21. National Center for Health Statistics. National Health and Nutrition Examination Survey, 2013–2014. Atlanta: Centers for Disease Control and Prevention; 2015. Accessed at www.cdc.gov/nchs/nhanes.htm [on 19 July 2016]
22. National Clinical Guideline Centre. Clinical Guideline 127. London: National Institute for Health and Clinical Excellence; 2011. Hypertension: The Clinical Management of Primary Hypertension in Adults.
23. Powers BJ, Oddone EZ, Bosworth HB. Measuring blood pressure for decision making and quality reporting. Reply to Wexler R [letter]. *Ann Intern Med.* 2011; 155:565.doi: 10.7326/0003-4819-155-8-201110180-00027
24. Kent DM, Rothwell PM, Ioannidis JP, Altman DG, Hayward RA. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. *Trials.* 2010; 11:85.doi: 10.1186/1745-6215-11-85 [PubMed: 20704705]
25. Dailey G, Wang E. A review of cardiovascular outcomes in the treatment of people with type 2 diabetes. *Diabetes Ther.* 2014; 5:385–402. DOI: 10.1007/s13300-014-0091-x [PubMed: 25515096]

Figure 1.

Model diagram of SMART to guide personalized treatment of BP.

BP = blood pressure; DBP = diastolic blood pressure; SBP = systolic blood pressure;

SMART = sequential, multiple assignment, randomized trial.

* Ethical trial practice is simulated at each stage by advancing persons with SBP >150 mm Hg or DBP >95 mm Hg to the next stage of treatment, which is consistent with current best practices. Treatment is also reduced to the next lower number of BP agents if a person experiences adverse effects of treatment, including hypotension.

Figure 2.

HTEs of BP therapy intensification on CVD events potentially hidden from detection in subgroup analyses in ACCORD and SPRINT.

For simplicity of presentation, the HTE for DBP is presented as binary, but it was modeled as a continuous function. (See the Supplement, available at Annals.org, for the full equations fitted to the results of both ACCORD and SPRINT; these equations also adjust for pretreatment and posttreatment systolic BP and pretreatment CVD risk.) Effects were estimated by simulating patients who met the trial inclusion criteria and plotting the distribution of anticipated HRs among them after calibration of the HR parameters to both ACCORD and SPRINT summary statistics. The CIs were estimated by repeating the simulation 10 000 times while sampling with replacement from the probability distributions of all input parameters (Supplement). ACCORD = Action to Control Cardiovascular Risk in Diabetes trial; BP = blood pressure; CVD = cardiovascular disease; DBP = diastolic blood pressure; HR = hazard ratio; HTE = heterogeneous treatment effect; SPRINT = Systolic Blood Pressure Intervention Trial.

Figure 3.

Estimated power of SMART versus a conventional treat-to-target trial to detect potential HTEs by sample size.

Results are based on the SMART study design shown in Figure 1 and a conventional treat-to-target design of a representative U.S study population with the following characteristics: age 50 to 75 y, 10-y Framingham CVD risk score 10%, SBP of 135 to 150 mm Hg and DBP of 65 to 100 mm Hg when receiving 1 or 2 BP agents, and no history of congestive heart failure. The specified HTEs for power calculations are those found in a microsimulation model that is fitted to both ACCORD and SPRINT results (see Supplement Figure 2 and equation 1 in the Supplement, both available at Annals.org). The dashed lines show where 2 of the lines (those pertaining to the number of agents and DBP in SMART) achieved 80% power. ACCORD = Action to Control Cardiovascular Risk in Diabetes trial; BP = blood pressure; CVD = cardiovascular disease; DBP = diastolic blood pressure; HTE = heterogeneous treatment effect; SBP = systolic blood pressure; SMART = sequential, multiple assignment, randomized trial; SPRINT = Systolic Blood Pressure Intervention Trial.

Table

Input Data From ACCORD and SPRINT Used to Construct the Microsimulation Models *

Variable	ACCORD (<i>n</i> = 4733) [†]	SPRINT (<i>n</i> = 9361) [‡]
Mean baseline SBP, <i>mm Hg</i>	139.2	139.7
Distribution of baseline SBP, %[§]		
132 mm Hg	NR	33.2
133–144 mm Hg	NR	33.1
145 mm Hg	NR	33.8
Mean baseline DBP, <i>mm Hg</i>	76.0	78.1
Framingham 10-y CVD risk score, %	NR	20.1
Framingham 10-y CVD risk score 15%, %	NR	61.4
Mean SBP at follow-up, <i>mm Hg</i>		
Standard group	133.5	134.6
Intensive group	119.3	121.5
Mean DBP at follow-up, <i>mm Hg</i>		
Standard group	70.5	76.3
Intensive group	64.4	68.7
Mean agents, <i>n</i>		
Standard group	2.1	1.8
Intensive group	3.4	2.7
Distribution of agents, %		
0 agents		
Standard group	7	11.3
Intensive group	1	2.7
1 agent		
Standard group	24	31.1
Intensive group	8	10.5
2 agents		
Standard group	30	33.3
Intensive group	20	30.5
3 agents		
Standard group	23	17.2

Variable	ACCORD (<i>n</i> = 4733) [†]	SPRINT (<i>n</i> = 9361) [‡]
Intensive group	30	31.8
4 agents		
Standard group	16	6.9
Intensive group	41	24.3

ACCORD = Action to Control Cardiovascular Risk in Diabetes trial; CVD = cardiovascular disease; DBP = diastolic blood pressure; NR = not reported; SBP = systolic blood pressure; SPRINT = Systolic Blood Pressure Intervention Trial.

^{*}We simulated the population of participants enrolled in each trial to match these baseline characteristics, then calibrated an equation (equation 1 in the Supplement, available at Annals.org) to capture how changes in SBP and DBP, baseline Framingham risk score, and total number of blood pressure agents predicted the hazard ratio of CVD outcomes (Figure 3).

[†]Reference 2.

[‡]Reference 3.

[§]These percentages do not sum to 100, probably due to rounding.

// ACCORD publications rounded these values to the nearest 1%, whereas SPRINT publications rounded them to the nearest 0.1%.