



# HHS Public Access

Author manuscript

*Lancet Diabetes Endocrinol.* Author manuscript; available in PMC 2018 February 16.

Published in final edited form as:

*Lancet Diabetes Endocrinol.* 2017 October ; 5(10): 808–815. doi:10.1016/S2213-8587(17)30176-6.

## Targeting weight loss interventions to reduce cardiovascular complications of type 2 diabetes: a machine learning-based post-hoc analysis of heterogeneous treatment effects in the Look AHEAD trial

Aaron Baum, PhD<sup>\*</sup>, Joseph Scarpa, PhD<sup>\*</sup>, Emilie Bruzelius, MPH, Ronald Tamler, MD, Sanjay Basu, MD, and James Faghmous, PhD

Department of Health System Design and Global Health, Arnhold Institute for Global Health (A Baum PhD, J Scarpa PhD, E Bruzelius MPH, J Faghmous PhD), Division of Endocrinology, Diabetes, and Bone Disease (R Tamler MD), Icahn School of Medicine at Mount Sinai, New York, NY, USA; Department of Epidemiology, Joseph L Mailman School of Public Health, Columbia University, New York, NY, USA (E Bruzelius); and Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA (S Basu MD)

### Summary

**Background**—The Action for Health in Diabetes (Look AHEAD) trial investigated whether long-term cardiovascular disease morbidity and mortality could be reduced through a weight loss intervention among people with type 2 diabetes. Despite finding no significant reduction in cardiovascular events on average, it is possible that some subpopulations might have derived benefit. In this post-hoc analysis, we test the hypothesis that the overall neutral average treatment effect in the trial masked important heterogeneous treatment effects (HTEs) from intensive weight loss interventions.

**Methods**—We used causal forest modelling, which identifies HTEs, using a random half of the trial data (the training set). We applied Cox proportional hazards models to test the potential HTEs on the remaining half of the data (the testing set). The analysis was deemed exempt from review by the Columbia University Institutional Review Board, Protocol ID# AAAO3003.

**Findings**—Between Aug 22, 2001, and April 30, 2004, 5145 patients with type 2 diabetes were enrolled in the Look AHEAD randomised controlled trial, of whom 4901 were included in the The National Institute of Diabetes and Digestive and Kidney Diseases Repository and included in our analyses: 2450 for model development and 2451 in the testing dataset. Baseline HbA<sub>1c</sub> and self-

---

Correspondence to: Aaron Baum, Department of Health System Design and Global Health, Arnhold Institute for Global Health, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA aaron.baum@mssm.edu.

<sup>\*</sup>These authors contributed equally

### Contributors

AB and JS had full access to all the data in the study and had final responsibility for the decision to submit for publication. AB, JS, and JF were responsible for the study concept and design. AB, JS, EB, SB, and JF did acquisition, analysis, and interpretation of data. AB, EB, SB, and JF drafted the manuscript. AB, JS, EB, RT, SB, and JF were responsible for critical revision of the manuscript for important intellectual content. AB and JF did statistical analysis. AB and JF supervised the study.

### Declaration of interests

The authors declare no competing interests.

reported general health distinguished participants who differentially benefited from the intervention. Cox models for the primary composite cardiovascular outcome revealed a number needed to treat of 28.9 to prevent 1 event over 9.6 years among participants with HbA<sub>1c</sub> 6.8% or higher, or both HbA<sub>1c</sub> less than 6.8% and Short Form Health Survey (SF-36) general health score of 48 or more (2101 [86%] of 2451 participants in the testing dataset; 167 [16%] of 1046 primary outcome events for intervention *vs* 205 [19%] of 1055 for control, absolute risk reduction of 3.46%, 95% CI 0.21–6.73%, *p*=0.038). By contrast, participants with HbA<sub>1c</sub> less than 6.8% and baseline SF-36 general health score of less than 48 (350 [14%] of 2451 participants in the testing data; 27 [16%] of 171 primary outcome events for intervention *vs* 15 [8%] of 179 primary outcome events for control) had an absolute risk increase of the primary outcome of 7.41% (0.60 to 14.22, *p*=0.003).

**Interpretation**—Look AHEAD participants with moderately or poorly controlled diabetes (HbA<sub>1c</sub> 6.8% or higher) and subjects with well controlled diabetes (HbA<sub>1c</sub> less than 6.8%) and good self-reported health (85% of the overall study population) averted cardiovascular events from a behavioural intervention aimed at weight loss. However, 15% of participants with well controlled diabetes and poor self-reported general health experienced negative effects that rendered the overall study outcome neutral. HbA<sub>1c</sub> and a short questionnaire on general health might identify people with type 2 diabetes likely to derive benefit from an intensive lifestyle intervention aimed at weight loss.

**Funding**—None.

---

## Introduction

Cardiovascular disease remains the leading cause of death among people with type 2 diabetes.<sup>1–3</sup> Short-term and non-randomised studies previously reported associations between weight loss among people with type 2 diabetes and improved cardiovascular disease risk factors or outcomes.<sup>4,5</sup> To assess whether long-term cardiovascular disease morbidity and mortality could be reduced through weight loss interventions, the Action for Health in Diabetes (Look AHEAD) trial randomised patients to either an intensive lifestyle intervention focused on weight loss achieved through healthy eating and increased physical activity (intervention group) or diabetes support and education (control group).<sup>6</sup> The study was stopped early due to a futility analysis, with no significant between-group differences in the primary composite outcome of first occurrence of death from cardiovascular causes, non-fatal myocardial infarction, non-fatal stroke, or hospitalisation for angina.<sup>6</sup> The study reported no significant between-group differences in pre-specified composite secondary outcomes, individual cardiovascular events, or interactions among the pre-specified subgroups.

As with many trials that have reported negative or neutral average treatment effects, statistical commentators have been concerned that the average study result could mask important heterogeneous treatment effects (HTEs), or systematically different outcomes among different types of study subjects.<sup>7</sup> Traditional subgroup analyses will typically fail to identify such HTEs, because they are underpowered and are susceptible to estimation bias and multiple testing errors. Additionally, subgroup analyses generally only consider one factor at a time, rather than combinations of factors that are typically thought to generate

HTEs.<sup>8</sup> Yet, detecting HTEs is crucial to practicing clinicians, since identifying individuals who might benefit from an intervention is required to avoid preventable complications of type 2 diabetes. Furthermore, both private and public health-care payers increasingly fund lifestyle interventions directed through clinical settings.<sup>9</sup> Ignoring HTEs might lead to lack of reimbursement for weight loss programmes, which would neglect potential benefits of such programmes for some populations.

To address the limitations of standard subgroup analyses, machine learning theorists devised the method of causal forest analysis<sup>10</sup> (appendix). Machine learning methods broadly aim to reveal new insights from data, without specifying a hypothesis a priori. These methods are employed in a wide range of tasks from speech recognition to autonomous vehicles and are increasingly applied to biomedical sciences for biomarker discovery, disease progression, and automated disease detection.<sup>11–13</sup> Causal forest analysis identifies subgroups by building numerous decision trees from prespecified covariates in a random subsample of the data and avoids multiple hypothesis testing by estimating model coefficients for subgroups defined by those covariate combinations on another subsample (honest estimation approach). The repeated data partitioning, internal cross-validation, and honest estimation approach minimises the risk of overfitting and produces unbiased HTE estimates that might be missed by standard subgroup analyses.<sup>10,14,15</sup>

We applied the causal forest method to fit Cox proportional hazards models to the Look AHEAD data. We tested the hypothesis that the overall neutral average treatment effect in the trial masked important HTEs from intensive weight loss interventions.

## Methods

### Study design and participants

Study design and reporting was based on the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement, a standardised, evidence-based set of recommendations for reporting prediction modelling studies.<sup>16</sup>

The study sample used for model development consisted of all participants in the Look AHEAD trial that were included in The National Institute of Diabetes and Digestive and Kidney Diseases Repository. Look AHEAD was a randomised, controlled, open-label trial of intensive lifestyle intervention focused on weight loss achieved through healthy eating and increased physical activity (intervention group) or diabetes support and education (control group) done at 16 clinical sites in the USA between Aug 22, 2001, and April 30, 2004.<sup>6</sup> Patients eligible for inclusion in the Look AHEAD trial were aged 45–75 years with a history of type 2 diabetes and overweight or obesity (BMI 25 kg/m<sup>2</sup> or higher, or 27 kg/m<sup>2</sup> or higher if taking insulin), blood pressure (BP) 160/100 mm Hg or less, HbA<sub>1c</sub> 11% or less, and plasma triglyceride density less than 600 mg/dL (6.8 mmol/L). Exclusion criteria included type 1 diabetes, adherence problems, and diseases limiting lifespan or affecting safety.

Eligible patients were randomly assigned to the intervention or control group. The lifestyle intervention included individual and group counselling sessions occurring weekly for the initial 6 months of the trial and decreasing gradually over the course of the trial. Participants in the control group received brief diet and exercise educational sessions and social support. Medical history, demographic, and social information were collected at baseline. Among all participants, weight loss, metabolic data, and medication use were obtained annually. Enrolled participants were queried about medical events and hospital admissions every 6 months.

## Outcomes

As per the Look AHEAD trial protocol, for our study we defined the primary outcome as the trial's primary composite cardiovascular outcome, which was the first occurrence of death from cardiovascular causes, non-fatal myocardial infarction (MI), non-fatal stroke, or hospitalisation for angina, and our three composite secondary outcomes as: cardiovascular mortality, non-fatal MI, or non-fatal stroke; all-cause mortality, non-fatal MI, or non-fatal stroke; and all-cause mortality, non-fatal MI, non-fatal stroke, hospital admission for angina, hospital admission for coronary artery bypass grafting, percutaneous coronary intervention, or heart failure, carotid endarterectomy, or peripheral vascular disease. We studied 84 baseline predictors from four major categories to estimate HTEs. These predictors included sociodemographic variables, medical history, laboratory values, and behavioural measures. All 84 predictors are provided in the appendix.

## Statistical analysis

We applied causal forest analysis to estimate subgroups with substantially different HTEs within a randomly sampled half of the overall trial population (the training dataset,  $n=2450$ , figure 1). We first constructed 1000 causal trees,<sup>10</sup> a type of decision tree built by repeated regression. Decision trees identify subgroups by producing a partition of the sample in which subgroups share similar predictions or classifications that are not limited by model specification assumptions. For each causal tree, 50% of the data is randomly selected without replacement as training data and the algorithm sequentially partitions this subset into covariate subgroups. For each level of the tree, the algorithm examines all possible split points for each covariate and selects covariate and split point pairs that minimise variation in the average treatment effect within each subgroup:

$$\Delta y = \bar{y}_{treated} - \bar{y}_{control}$$

To generate testable HTE hypotheses from the results of the forest, we developed a heuristic to select the subgroups (leaves) most representative of the treatment effect heterogeneity identified by the forest (appendix). For these most representative subgroups, we used Cox proportional hazards regressions to calculate hazard ratios, 95% CIs, and likelihood-ratio tests to estimate the significance of differences in hazard rates for the primary and secondary outcomes between the groups. Following standardised protocols for detection of HTEs, the Cox models contain terms for study group assignment, a subgroup dummy variable, and their interaction.<sup>17</sup> We did robustness checks of the statistical significance of our findings

through bootstrapping. Additionally, to explore potential mechanisms through which differential intervention response across subgroups might have operated, we estimated the significance of differences in intermediate health outcomes and intervention process indicators within each subgroup (appendix). Data from the Look AHEAD Trial were obtained from The National Institute of Diabetes and Digestive and Kidney Diseases Repository. All analyses were performed in *R* (version 3.2.2, *R* Foundation for Statistical Computing, Vienna) and Stata (version 14, Stata Corp, College Station, Texas). The analysis was deemed exempt from review by the Columbia University Institutional Review Board, Protocol ID# AAAO3003.

### Role of the funding source

There was no funding source for this study. AB and JS had full access to all of the data and the final responsibility to submit for publication.

## Results

The trial recruited patients from Aug 22, 2001, to April 30, 2004. The study sample for model development included 2450 (47.6%) of 5145 randomly assigned Look AHEAD trial participants. 244 (4.7%) of 5145 Look AHEAD participants were omitted in the National Institute of Diabetes and Digestive and Kidney Diseases Repository dataset, and 2451 (47.6%) of 5145 participants were omitted from model development to preserve for internal testing (figure 1). The training sample (n=2450) included 1231 (47.8%) of 2570 participants in the intervention group, and 1219 (47.4%) of 2575 participants in the control group. Characteristics of the included Look AHEAD study sample are provided in table 1; the included participant sample for model development averaged 58.9 years old, was 59% female, and had an average BMI of 36.0 kg/m<sup>2</sup>. Participants were followed up for a mean of 8.5 years. In the training sample, a total of 199 (16.2%) of 1231 included participants from the intervention group experienced a primary outcome event, and 186 (15.2%) of 1219 included participants from the control group experienced a primary outcome event.

The causal forest model revealed two covariates, baseline HbA<sub>1c</sub> and general health (as self-reported on the SF-36 health survey),<sup>18</sup> were of primary importance in distinguishing individuals with high versus low benefit from the intensive weight loss intervention (appendix). A third variable of importance, also reported on the SF-36, was self-reported mental health, which was highly associated to SF-36 general health in the cohort (rho 0.41, p<0.0001). We note that age, sex, and ethnicity explained a minimal portion of the variance in treatment effect, ranking as the 36th, 59th, and 48th most important forest variables, respectively.

The causal forest model partitioned the trial participants into six subgroups (leaves), of which two leaves included less than 10% of the study sample and were therefore excluded. We refer to the remaining four subgroups as subgroup 1 (baseline HbA<sub>1c</sub> less than 6.8% and baseline SF-36 general health score less than 48); subgroup 2 (baseline HbA<sub>1c</sub> less than 6.8% and baseline SF-36 general health score 48 or more); subgroup 3 (baseline HbA<sub>1c</sub> 6.8% or higher and baseline SF-36 Mental Component Summary [MCS] score less than 54); and subgroup 4 (baseline HbA<sub>1c</sub> 6.8% or higher and baseline SF-36 MCS score 54 or

higher). Of the four subgroups, only subgroup 1 and subgroup 2 were defined by covariates identified as primary in the forest (figure 2). We focused our analysis on these two subgroups, which contained 16% and 24% of the training data, respectively.

The study sample for testing of the HTE hypotheses (testing set) included 2451 (47.6%) of 5145 randomly assigned Look AHEAD trial participants (figure 1). The proportion of patients in the control and intervention groups of the testing data was nearly equivalent to those proportions in the training data, ensuring that the data split was not biased. Specifically, the testing sample included 1217 (47.3%) of 2570 participants in the intervention group, and 1234 (48.0%) of 2575 participants in the control group. Analyses of the other participant characteristics (table 1) validate that covariates were balanced across training and testing data, and between control and intervention groups of each data subset (appendix).

Using the testing dataset, those not in subgroup 1, ie, participants with HbA<sub>1c</sub> 6.8% or higher or both HbA<sub>1c</sub> less than 6.8% and SF-36 general health score 48 or more (4152 [84.7%] of 4901 overall trial participants, 2051 [83.7%] of 2450 participants in the model development dataset, and 2101 [85.7%] of 2451 participants in the testing dataset), revealed a number needed to treat (NNT) of 28.9 to prevent one primary outcome event over 9.6 years (16.0% events for intervention vs 19.4% for control, absolute risk reduction [ARR] 3.46%; 95% CI 0.21 to 6.73; p=0.038, table 2 and figure 3). For the first secondary outcome, this subgroup had an NNT of 38.9 to prevent one event over 9.6 years (10.1% events for intervention vs 12.7% for control, ARR 2.57%, 95% CI -0.15 to 5.28; p=0.064, table 2). For the second secondary outcome, this subgroup had an NNT of 26.8 to prevent one event over 9.6 years (19.2% events for intervention vs 22.9% for control; ARR 3.72%, 95% CI 0.24 to 7.21; p=0.037, table 2). For the third secondary outcome this subgroup had a non-significant NNT of 40.6 to prevent one event over 9.6 years (22.6% events for intervention vs 25.0% for control, ARR 2.46%, 95% CI -1.18 to 6.10; p=0.185, table 2).

By contrast, the remaining subgroup of participants with HbA<sub>1c</sub> less than 6.8% and SF-36 general health scores less than 48 (subgroup 1; 750 [15.3%] of 4901 overall trial participants, 400 [16.3%] of 2450 participants in the model development dataset and 350 [14.3%] of 2451 participants in the testing dataset) had an absolute risk increase of the primary outcome of 7.41% (95% CI 0.60 to 14.22, p=0.033). For the first secondary outcome, this subgroup had a non-significant increase in absolute risk of 3.80% (-1.90 to 9.50, p=0.191). For the second secondary outcome, this subgroup had an absolute increase in risk of 8.15% (0.53 to 15.77, p=0.036). For the third secondary outcome this subgroup had a non-significant absolute increase in risk of 5.97% (-2.08 to 14.20, p=0.15; table 2).

In exploratory analyses, we found evidence of greater intermediate improvement in HbA<sub>1c</sub>, self-reported mental health, and blood pressure due to the intervention among those not in subgroup 1 compared with those in subgroup 1 (appendix). By contrast, participants in subgroup 1 reported fewer minutes of exercise in the first 6 months (-495 minutes; p<0.0009) and last 6 months (-924 minutes; p<0.0009) of the intervention year compared with those not in subgroup 1 (appendix).

Additionally, using the testing dataset, participants in subgroup 2, ie, with HbA<sub>1c</sub> less than 6.8% and SF-36 general health score 48 or higher (1060 [21.6%] of 4901 overall trial participants, 528 [21.5%] of 2450 participants in the model development dataset and 532 [21.7%] 2451 in the testing dataset), had an NNT of 12.2 to prevent 1 primary outcome event over 9.6 years (10.8% events for intervention vs 19.0% for control, ARR 8.22%, 95% CI 2.17 to 14.27; p=0.007, table 2 and figure 3). By contrast, the remaining subgroup of participants with HbA<sub>1c</sub> 6.8% or higher or both HbA<sub>1c</sub> less than 6.8% and SF-36 general health score lower than 48 (3841 [78.3%] of 4190 overall trial participants, 922 [78.5%] of 2450 in the model development dataset and 1919 [78.3%] of 2451 in the testing dataset) had a non-significant ARR in the primary outcome of 0.05% (95% CI -3.35 to 3.45; p=0.977, table 2). Similar to the original Look AHEAD trial report (hazard ratio in the intervention group, 0.95; 95% CI 0.83 to 1.09; p=0.51),<sup>6</sup> in the testing dataset there was no significant difference in the primary outcome between the intervention and control groups in the overall sample of testing data (1.84 events per 100 person-years in the intervention group and 2.12 events per 100 person-years in the control group; hazard ratio in the intervention group 0.87; 95% CI 0.71 to 1.05; p=0.15).

## Discussion

In this analysis of the Look AHEAD trial, we identified a subgroup that experienced reduced cardiovascular events after a behavioural intervention aimed at weight loss. Using a machine learning method, called causal forest, on a training set of trial data to identify HTEs, then applying Cox proportional hazards on the testing set of trial data, we found that 85% of the study population averted cardiovascular events after the intervention; this subgroup comprised participants with moderately or poorly controlled diabetes (HbA<sub>1c</sub> 6.8% or higher) at baseline, and participants with both well controlled diabetes (HbA<sub>1c</sub> less than 6.8%) and good self-reported health at baseline. The overall Look AHEAD study outcome was rendered neutral due to the 15% of participants with well controlled diabetes and poor self-reported general health who experienced negative effects from the intervention.

The Diabetes Prevention Program established the usefulness of lifestyle changes for patients at risk for type 2 diabetes,<sup>19</sup> leading to a paradigm shift that values systematic behavioural programmes as medical interventions. Recommending weight loss through lifestyle intervention is considered best practice for patients with type 2 diabetes and is recommended by the American Diabetes Association.<sup>20</sup> While Look AHEAD failed to achieve its primary objective of cardiovascular risk reduction through a lifestyle programme aimed at weight reduction, commenters have highlighted the manifold beneficial effects of the intervention<sup>21</sup> and lamented that the study duration was too short to thoroughly assess diabetic complications.<sup>22</sup> Our finding that the outcome would have been positive for 85% of participants supports this argument. However, our study provides what could be the first suggestive evidence of an adverse reaction to what is generally considered a common-sense and innocuous intervention: 15% of subjects had substantially increased risk for the primary outcome (interaction term p=0.006), rendering the overall Look AHEAD study outcome neutral. These at-risk participants (subgroup 1) had baseline mild or well treated diabetes (HbA<sub>1c</sub> less than 6.8%) and baseline negative perception of their health status (SF-36 general health score less than 48); of note, SF-36 general health score was a strong correlate

of self-reported mental health ( $\rho$  0.41,  $p < 0.0001$ ), and Beck's depression score ( $\rho$  -0.42,  $p < 0.0001$ ). This finding supports prior studies that suggest that psychosocial factors might influence the efficacy of lifestyle interventions.<sup>23</sup> Indeed, our exploratory analysis suggests that intervention compliance was substantially poorer among subgroup 1. This is consistent with a literature documenting the importance of assessing patients' readiness for change when recommending behavioural interventions.<sup>24</sup> The observation that subgroup 1 (ie, baseline HbA<sub>1c</sub> less than 6.8% and baseline SF-36 general health score less than 48) had little or no improvement in several intermediate health outcomes also complements the Look AHEAD investigators' post-hoc analyses showing that participants who lost less than 10% of their bodyweight had greater risk of the primary outcome.<sup>25</sup>

Our analysis has several important limitations. We explicitly mitigated multiple testing concerns by partitioning the data into independent subsets. While this approach retains the validity of inference, it sacrifices statistical power. Given the low number of primary outcome events in the cohort, sample splitting might have limited our ability to identify subgroups with relatively small (positive or negative) effect sizes. Thus, the method's use is limited to randomised controlled trials with enough initial power to support inference on a partition. Our method also uses an intuitive heuristic to identify a subgroup that reflects the forest's average output, but a standard statistical approach to determining a representative subgroup does not exist and would be preferable. Additionally, because the two subgroups we tested are defined by the same covariates, the two tests for heterogeneity are not independent. Finally, we did not account for unequal observation periods while identifying the subgroups in the training set because the causal forest algorithm does not inherently account for censoring. To address this limitation of causal forest analysis, we used Cox proportional hazards models while estimating HTEs in the preserved data. Enhancing the machine learning algorithm to compare subgroup treatment effects on a relative, rather than absolute, scale is an important area for future research.

Further prospective investigation of our findings is necessary. To test our finding of potential benefit for a subgroup of patients with type 2 diabetes, prospective assessment of people with type 2 diabetes who meet the criteria detected here could be performed.<sup>26-28</sup> Even before such a study, our investigation illustrates that advances in machine learning for causal inference can identify important HTEs hidden among large subgroups within existing trials, even trials that report average negative effects. Our findings suggest that data-driven methods for HTE hypothesis generation can reveal otherwise undiscovered and clinically meaningful relationships between interventions, outcomes, and subgroups, and can complement expert-based preregistered subgroup hypotheses. Identifying robust subgroup treatment effects can increase the quantity of clinically relevant findings generated by clinical trials and enable clinicians to better individualise patient care.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.



## Acknowledgments

JS is supported by the National Institute of Mental Health of the NIH under Award Number F30MH106293. SB is supported by grants from the National Institute on Minority Health and Health Disparities of the National Institutes of Health under Award Numbers DP2MD010478 and U54MD010724; and the National Heart, Lung, And Blood Institute of the National Institutes of Health under Award Number K08HL121056. EB and JF are supported by the National Science Foundation under Award Number 1464297.

## References

1. Tancredi M, Rosengren A, Svensson AM, et al. Excess mortality among persons with type 2 diabetes. *N Engl J Med*. 2015; 373:1720–32. [PubMed: 26510021]
2. Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HA. 10-year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med*. 2008; 359:1577–89. [PubMed: 18784090]
3. Seshasai SR, Kaptoge S, et al. Emerging Risk Factors C. Diabetes mellitus, fasting glucose, and risk of cause-specific death. *N Engl J Med*. 2011; 364:829–41. [PubMed: 21366474]
4. National Institutes of Health. Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults—the evidence report. *Obes Res*. 1998; 6(suppl 2):51S–209S. [PubMed: 9813653]
5. Harrington M, Gibson S, Cottrell RC. A review and meta-analysis of the effect of weight loss on all-cause mortality risk. *Nutr Res Rev*. 2009; 22:93–108. [PubMed: 19555520]
6. Wing RR, Bolin P, et al. Look AHEAD Research Group. Cardiovascular effects of intensive lifestyle intervention in type 2 diabetes. *N Engl J Med*. 2013; 369:145–54. [PubMed: 23796131]
7. Basu S, Sussman JB, Hayward RA. Detecting heterogeneous treatment effects to guide personalized blood pressure treatment: a modeling study of randomized clinical trials. *Ann Intern Med*. 2017; 166:354–60. [PubMed: 28055048]
8. VanderWeele TJ, Knol MJ. Interpretation of subgroup analyses in randomized trials: heterogeneity versus secondary interventions. *Ann Intern Med*. 2011; 154:680–83. [PubMed: 21576536]
9. Kaiser Family Foundation. [accessed Feb 4, 2017] Preventive services covered by private health plans under the Affordable Care Act. 2015. <http://kff.org/health-reform/fact-sheet/preventive-services-covered-by-private-health-plans/>
10. Athey S, Imbens G. Recursive partitioning for heterogeneous causal effects. *Proc Natl Acad Sci USA*. 2016; 113:7353–60. [PubMed: 27382149]
11. Wong TY, Bressler NM. Artificial intelligence with deep learning technology looks into diabetic retinopathy screening. *JAMA*. 2016; 316:2366–67. [PubMed: 27898977]
12. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine learning and data mining methods in diabetes research. *Comput Struct Biotechnol J*. 2017
13. Razavian N, Blecker S, Schmidt AM, Smith-McLallen A, Nigam S, Sontag D. Population-level prediction of type 2 diabetes from claims data and analysis of risk factors. *Big Data*. 2015; 3:277–87. [PubMed: 27441408]
14. Wagner, S. [accessed Feb 4, 2017] Asymptotic theory for random forests. 2014. <https://arxiv.org/pdf/1405.0352v2.pdf>
15. Athey, S., Imbens, G. [accessed Feb 4, 2017] Recursive partitioning for heterogeneous causal effects. 2015. <https://arxiv.org/abs/1504.01132v3>
16. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD). *Ann Intern Med*. 2015; 162:735–36.
17. 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. [PubMed: 20704705]
18. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992; 30:473–83. [PubMed: 1593914]
19. Knowler WC, Barrett-Connor E, Fowler SE, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med*. 2002; 346:393–403. [PubMed: 11832527]

20. Colberg SR, Sigal RJ, Yardley JE, et al. Physical activity/exercise and diabetes: A position statement of the American Diabetes Association. *Diabetes Care*. 2016; 39:2065–79. [PubMed: 27926890]
21. Pi-Sunyer X. The Look AHEAD trial: a review and discussion of its outcomes. *Curr Nutr Rep*. 2014; 3:387–91. [PubMed: 25729633]
22. Bennett PH. The Look AHEAD study: a missed opportunity. *Lancet Diabetes Endocrinol*. 2:775–76.
23. Gonzalez JS, Peyrot M, McCarl LA, et al. Depression and diabetes treatment nonadherence: a meta-analysis. *Diabetes Care*. 2008; 31:2398–403. [PubMed: 19033420]
24. Delahanty LM, Conroy MB, Nathan DM. Psychological predictors of physical activity in the diabetes prevention program. *J Am Diet Assoc*. 2006; 106:698–705. [PubMed: 16647327]
25. Association of the magnitude of weight loss and changes in physical fitness with long-term cardiovascular disease outcomes in overweight or obese people with type 2 diabetes: a post-hoc analysis of the Look AHEAD randomised clinical trial. *Lancet Diabetes Endocrinol*. 4:91–21.
26. Fiore LD, Lavori PW. Integrating randomized comparative effectiveness research with patient care. *N Engl J Med*. 2016; 374:2152–58. [PubMed: 27248620]
27. Patel A, Webster R. Pragmatic trials for noncommunicable diseases: relieving constraints. *PLoS Med*. 2016; 13:e1001986. [PubMed: 27022969]
28. Staa TP, Goldacre B, Gulliford M, et al. Pragmatic randomised trials using routine electronic health records: putting them to the test. *BMJ*. 2012; 344:e55. [PubMed: 22315246]

## Research in context

### Evidence before this study

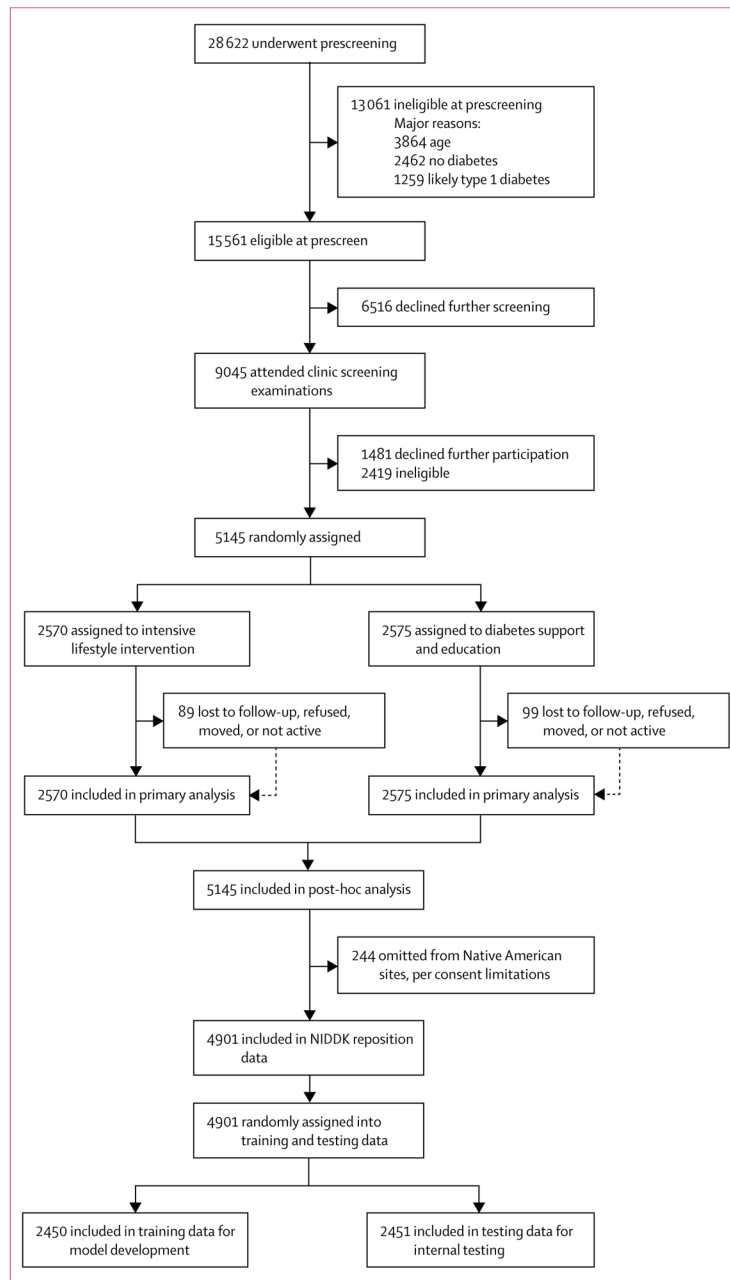
We searched PubMed for studies published between Jan 1, 2006, and Dec 31, 2016, using the terms “cardiovascular disease”, “weight loss” and “diabetes mellitus”. Results of meta-analyses suggest that type 2 diabetes confers an excess risk for coronary heart disease and prior randomised and non-randomised controlled trials show that intensive weight loss interventions can improve cardiovascular risk factors in this population. However, the Look AHEAD study, a landmark randomised controlled trial investigating whether long-term cardiovascular disease morbidity and mortality could be reduced through a weight loss intervention among people with type 2 diabetes, reported no significant benefit with respect to primary or secondary outcomes or to the individual cardiovascular events making up the composite outcomes. However, the overall neutral average treatment effect in the trial may have masked important heterogeneous treatment effects of the intensive weight loss intervention among subpopulations.

### Added value of this study

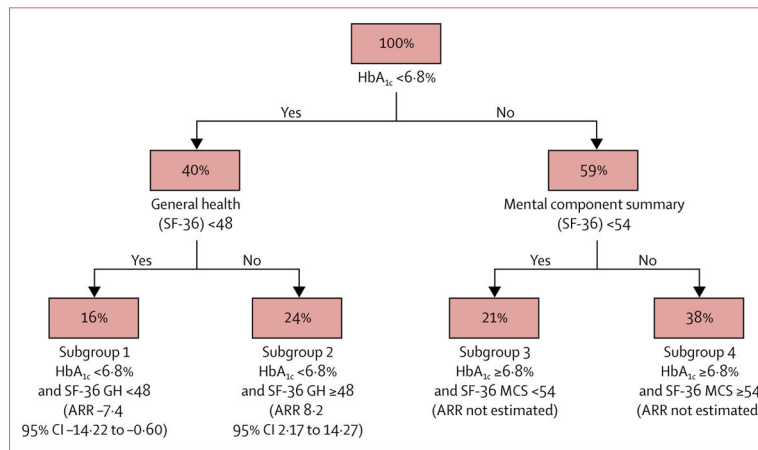
We show that those participants in the Look AHEAD trial who had HbA<sub>1c</sub> 6·8% or more, or both HbA<sub>1c</sub> less than 6·8% and above average self-reported general health, experienced a clinically meaningful, significant reduction in cardiovascular events (the composite primary outcome) from the intensive weight loss intervention, despite the overall null trial findings. These participants constituted 85% of the overall study sample, but were counterbalanced by another 15% with moderate HbA<sub>1c</sub> levels (less than 6·8%) and poor self-reported general health who experienced negative effects that rendered the overall study outcome neutral.

### Implications of all the available evidence

The findings suggest that HbA<sub>1c</sub> and a short questionnaire on general health may identify persons with type 2 diabetes likely to benefit from an intensive lifestyle intervention to avert cardiovascular events. More broadly, our investigation demonstrates that recent advances in machine learning for causal inference can reveal important heterogeneity hidden among large subgroups within existing trials.

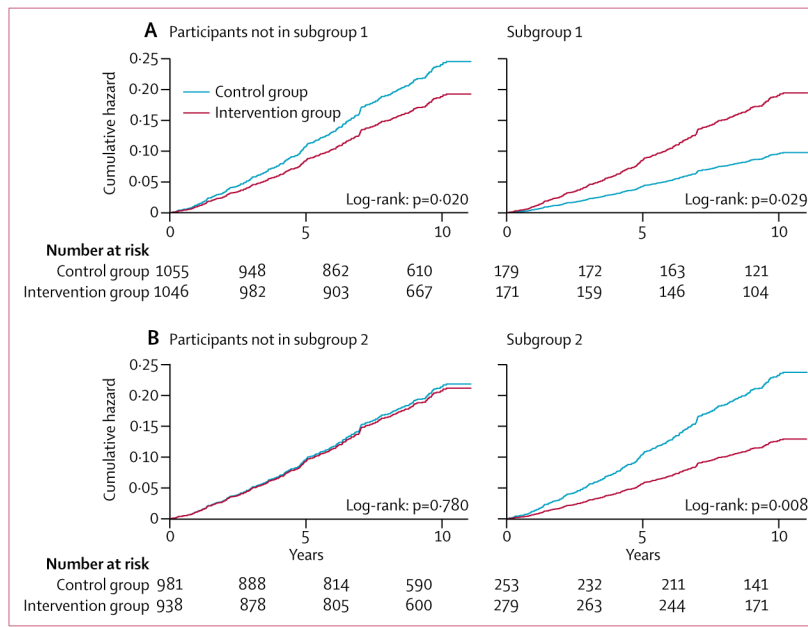


**Figure 1. Trial profile**  
 NIDDK=National Institute of Diabetes and Digestive and Kidney Diseases.



**Figure 2. Subgroups identified by the representative causal tree**

Nodes indicate the percent of the training data sample in each subdivision of the data, with the covariate and split point identified underneath. For example, subgroup 1 contains 16% of the training data and includes participants with baseline HbA<sub>1C</sub> less than 6.8% and baseline SF-36 general health score less than 48, and subgroup 2 contains 24% of the training data and includes participants with baseline HbA<sub>1C</sub> less than 6.8% and baseline SF-36 general health score 48 or higher. Absolute risk reduction refers to the primary outcome of the Look AHEAD trial. For clarity, we removed from this figure two subgroups each containing 1% of the training data. SF-36 Short Form Health Survey. GH=general health. MCS=Mental Component Summary. ARR=absolute risk reduction (calculated using the testing data).



**Figure 3. Cumulative hazard curves for the primary composite endpoint**  
 Cumulative hazard curves across treated and control groups are shown for (A) participants not in (left) and in (right) subgroup 1 (baseline HbA<sub>1C</sub> <6.8% and SF-36 general health score <48); and (B) participants not in (left) and in (right) subgroup 2 (baseline HbA<sub>1C</sub> <6.8% and SF-36 general health score ≥ 48). The primary outcome was a composite of death from cardiovascular causes, non-fatal myocardial infarction, non-fatal stroke, or hospitalisation for angina.

**Table 1** Baseline characteristics of patients across treated and control groups in the overall dataset, the training subset, and the testing subset

	Overall (n=4901)			Training data (n=2450)			Testing data (n=2451)		
	Treated (n=2448)	Control (n=2453)	p value	Treated (n=1231)	Control (n=1219)	p value	Treated (n=1217)	Control (n=1234)	p value
Age (years)	58.8 (6.7)	59.1 (6.8)	0.08	58.7 (6.8)	59.2 (6.9)	0.05	58.8 (6.7)	59.0 (6.7)	0.62
Sex									
Male	1004 (41%)	1006 (41%)	1.00	505 (41%)	505 (41%)	0.84	499 (41%)	506 (41%)	0.84
Female	1444 (59%)	1447 (59%)	1.00	726 (59%)	719 (59%)	0.84	718 (59%)	728 (59%)	0.84
Race*	3.20 (1.19)	3.19 (1.19)	0.98	3.19 (1.20)	3.15 (1.22)	0.39	3.20 (1.19)	3.24 (1.16)	0.37
History of cardiovascular disease (0=no; 1=yes)	1.15 (0.35)	1.14 (0.35)	0.40	1.15 (0.36)	1.14 (0.35)	0.38	1.14 (0.35)	1.14 (0.34)	0.75
History of smoking <sup>‡</sup>	2.55 (0.58)	2.53 (0.58)	0.35	2.56 (0.59)	2.53 (0.58)	0.10	2.54 (0.58)	2.54 (0.58)	0.73
Diabetes duration (years)	6.7 (6.5)	6.8 (6.3)	0.71	6.5 (6.1)	6.6 (6.2)	0.77	6.9 (6.9)	7.0 (6.4)	0.84
Weight (kg)	101.00 (19.7)	101.2 (18.9)	0.65	101.0 (19.5)	101.3	0.74	101.0 (19.8)	101.2 (19.2)	0.76
BMI kg/m <sup>2</sup>	35.9 (6.0)	36.0 (5.8)	0.67	36.0 (6.0)	36.0 (5.6)	0.93	35.9 (6.1)	36.0 (5.9)	0.49
HbA <sub>1c</sub> (%)	7.23 (1.13)	7.29 (1.19)	0.05	7.22 (1.11)	7.27 (1.19)	0.30	7.24 (1.15)	7.32 (1.19)	0.08
Systolic blood pressure (mm Hg)	128.4 (17.2)	129.7 (17.0)	0.01	128.4 (17.4)	129.6 (16.9)	0.05	128.5 (17.1)	129.8 (17.2)	0.07
HDL (mmol/L)	1.12 (0.31)	1.13 (0.31)	0.80	1.12 (0.30)	1.13 (0.31)	0.65	1.13 (0.32)	1.12 (0.30)	0.92
LDL (mmol/L)	2.90 (0.84)	2.91 (0.83)	0.87	2.89 (0.84)	2.92 (0.86)	0.47	2.92 (0.83)	2.90 (0.81)	0.61
Triglyceride (mg/dL)	2.06 (1.30)	2.027 (1.31)	0.38	2.03 (1.27)	2.03 (1.24)	0.95	2.09 (1.33)	2.02 (1.38)	0.25

Data are mean (SD).

\* 1=African American/Black (not Hispanic); 2=Hispanic; 3=Other/Mixed; 4=White.

<sup>‡</sup> 1=Never; 2=Past; 3=Present; 4=Missing.

**Table 2**

Observed outcomes by treatment group, stratified by subgroups identified by training data

	Number of patients treated with the intervention (number of events)	Number of control patients (number of events)	Absolute risk reduction, % (95% CI)	p value	Hazard ratio (95% CI)	p value
<b>Primary outcome: Death from cardiovascular causes, non-fatal myocardial infarction, non-fatal stroke, or hospitalisation for angina</b>						
Overall	1217 (194)	1234 (220)	1.89 (-1.08 to 4.85)	0.212	0.87 (0.71-1.05)	0.15
Subgroup 1	..	..	..	..	..	0.006 (interaction)
Baseline HbA <sub>1c</sub> less than 6.8% and baseline SF-36 general health score less than 48	171 (27)	179 (15)	-7.41 (-14.22 to -0.60)	0.033	1.99 (1.06-3.75)	0.03
Remainder of trial population	1046 (167)	1055 (205)	3.46 (0.21 to 6.73)	0.038	0.78 (0.63-0.96)	0.02
Subgroup 2	..	..	..	..	..	0.025 (interaction)
Baseline HbA <sub>1c</sub> less than 6.8% and baseline SF-36 general health score 48 or higher	279 (30)	253 (48)	8.22 (2.17 to 14.27)	0.007	0.55 (0.35-0.86)	0.01
Remainder of trial population	938 (164)	981 (172)	0.05 (-3.35 to 3.45)	0.977	0.97 (0.78-1.20)	0.78
<b>Secondary outcome 1: Death from cardiovascular causes, non-fatal myocardial infarction, or non-fatal stroke</b>						
Overall	1217 (123)	1234 (145)	1.64 (-0.83 to 4.11)	0.192	0.84 (0.66-1.07)	0.148
Subgroup 1	..	..	..	..	..	0.058 (interaction)
Baseline HbA <sub>1c</sub> less than 6.8% and baseline SF-36 general health score less than 48	171 (16)	179 (11)	-3.80 (-9.50 to 1.90)	0.191	1.67 (0.79-3.57)	0.182
Remainder of trial population	1046 (106)	1055 (134)	2.57 (-0.15 to 5.28)	0.064	0.77 (0.60-0.99)	0.044
Subgroup 2	..	..	..	..	..	0.058 (interaction)
Baseline HbA <sub>1c</sub> less than 6.8% and baseline SF-36 general health score 48 or higher	279 (21)	253 (35)	6.31 (1.05 to 11.57)	0.018	0.53 (0.31-0.91)	0.021
Remainder of trial population	938 (102)	981 (110)	0.34 (-2.47 to 3.14)	0.813	0.94 (0.72-1.24)	0.680
<b>Secondary outcome 2: Death from any cause, non-fatal myocardial infarction, non-fatal stroke, or hospitalisation for angina</b>						
Overall	1217 (235)	1234 (263)	2.00 (-1.18 to 5.19)	0.218	0.88 (0.74-1.05)	0.148
Subgroup 1	..	..	..	..	..	0.006 (interaction)



	Number of patients treated with the intervention (number of events)	Number of control patients (number of events)	Absolute risk reduction, % (95% CI)	p value	Hazard ratio (95% CI)	p value
Baseline HbA <sub>1c</sub> less than 6.8% and baseline SF-36 general health score less than 48	171 (34)	179 (21)	-8.15 (-15.77 to -0.53)	0.036	1.80 (1.05-3.10)	0.033
Remainder of trial population	1046 (201)	1055 (242)	3.72 (0.24 to 7.21)	0.037	0.080 (0.66-0.96)	0.019
Subgroup 2	..	..	..	..	..	0.015 (interaction)
Baseline HbA <sub>1c</sub> less than 6.8% and baseline SF-36 general health score 48 or higher	279 (38)	253 (59)	9.70 (3.11 to 16.29)	0.004	0.56 (0.37-0.85)	0.006
Remainder of trial population	938 (197)	981 (204)	-0.21 (-3.85 to 3.43)	0.911	0.98 (0.81-1.19)	0.848
<b>Secondary outcome 3: Death from any cause, non-fatal myocardial infarction, non-fatal stroke, hospitalisation for angina, CABG, PCI, hospital admission for heart failure, carotid endarterectomy, or peripheral vascular disease</b>						
Overall	1217 (272)	1234 (291)	1.23 (-2.10 to 4.46)	0.469	0.92 (0.78-1.08)	0.313
Subgroup 1	..	..	..	..	..	0.050 (interaction)
Baseline HbA <sub>1c</sub> less than 6.8% and baseline SF-36 general health score less than 48	171 (36)	179 (27)	-5.97 (-14.20 to 2.08)	0.146	1.47 (0.89-2.41)	0.129
Remainder of trial population	1046 (236)	1055 (264)	2.46 (-1.18 to 6.10)	0.185	0.86 (0.72-1.03)	0.097
Subgroup 2	..	..	..	..	..	0.060 (interaction)
Baseline HbA <sub>1c</sub> less than 6.8% and baseline SF-36 general health score 48 or higher	279 (50)	253 (65)	7.77 (0.75 to 14.79)	0.030	0.67 (0.47-0.97)	0.035
Remainder of trial population	938 (222)	981 (226)	-0.63 (-4.42 to 3.16)	0.745	1.00 (0.83-1.20)	0.976

An absolute risk reduction greater than 1 indicates the risk of an outcome was lower for the intervention group than the control group. A hazard ratio less than 1 indicates the risk of an outcome was lower for the intervention group than the control group. CABG=coronary artery bypass graft. PCI=percutaneous coronary intervention.