Characteristics Associated With Decreased or Increased Mortality Risk From Glycemic Therapy Among Patients With Type 2 Diabetes and High Cardiovascular Risk: Machine Learning Analysis of the ACCORD Trial

Sanjay Basu,[1,2] Sridharan Raghavan,[3,4] Deborah J. Wexler,[2,5] and Seth A. Berkowitz[2,5,6]

OBJECTIVE

Identifying patients who may experience decreased or increased mortality risk from intensive glycemic therapy for type 2 diabetes remains an important clinical challenge. We sought to identify characteristics of patients at high cardiovascular risk with decreased or increased mortality risk from glycemic therapy for type 2 diabetes using new methods to identify complex combinations of treatment effect modifiers.

RESEARCH DESIGN AND METHODS

The machine learning method of gradient forest analysis was applied to understand the variation in all-cause mortality within the Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial (N = 10,251), whose participants were 40–79 years old with type 2 diabetes, hemoglobin A_{1c} (HbA_{1c}) ≥7.5% (58 mmol/mol), cardiovascular disease (CVD) or multiple CVD risk factors, and randomized to target HbA_{1c} <6.0% (42 mmol/mol; intensive) or 7.0–7.9% (53–63 mmol/mol; standard). Covariates included demographics, BMI, hemoglobin glycosylation index (HGI; observed minus expected HbA_{1c} derived from prerandomization fasting plasma glucose), other biomarkers, history, and medications.

RESULTS

The analysis identified four groups defined by age, BMI, and HGI with varied risk for mortality under intensive glycemic therapy. The lowest risk group (HGI <0.44, BMI <30 kg/m^2, age <61 years) had an absolute mortality risk decrease of 2.3% attributable to intensive therapy (95% CI 0.2 to 4.5, P = 0.038; number needed to treat: 43), whereas the highest risk group (HGI ≥0.44) had an absolute mortality risk increase of 3.7% attributable to intensive therapy (95% CI 1.5 to 6.0; P < 0.001; number needed to harm: 27).

CONCLUSIONS

Age, BMI, and HGI may help individualize prediction of the benefit and harm from intensive glycemic therapy.

[1]Center for Primary Care and Outcomes Research, Center for Population Health Sciences, Departments of Medicine and Health Research and Policy, Stanford University, Palo Alto, CA
[2]Harvard Medical School, Boston, MA
[3]Department of Veterans Affairs Eastern Colorado Healthcare System, Denver, CO
[4]Division of General Internal Medicine, Department of Medicine, University of Colorado School of Medicine, Denver, CO
[5]Diabetes Unit, Massachusetts General Hospital, Boston, MA
[6]Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA

Corresponding author: Sanjay Basu, basus@stanford.edu.

Individualizing the glycemic target for patients with type 2 diabetes is now the guideline-recommended strategy (1), but how best to individualize glycemic targets remains unclear. A major reason for caution regarding intensive glycemic targets is the Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial (N = 10,251, conducted 2001–2009) (2), which was halted due to increased all-cause mortality in the intensive therapy arm. ACCORD targeted nearly normal glycemic levels in the intensive glycemic therapy arm, achieving a median hemoglobin A_{1c} (HbA_{1c}) of 6.4% (46 mmol/mol), compared with an achieved HbA_{1c} of 7.5% (58 mmol/mol) in the standard therapy arm. Meta-analyses of data from ACCORD and other trials find that microvascular events are reduced with intensive glycemic control (3), but the lack of overall mortality benefit in trials as well as the increased mortality observed in ACCORD renders uncertain the risk-to-benefit calculation in any given patient.

Although current guidelines do not recommend targets as low as those used in ACCORD, real-world evidence suggests many patients are treated with multidrug regimens to levels achieved in the intensive therapy arm of ACCORD (4–6). Therefore, understanding the heterogeneous treatment effects (HTEs) of intensive glycemic therapy with regard to mortality is important. Specifically, which subgroup of patients in ACCORD was most likely to experience increased mortality? Conversely, because some patients do derive cardiovascular benefit from glycemic therapy (7), did any subgroups in ACCORD experience benefit? Unfortunately, univariable subgroup analyses of the trial data have been unable to explain the major variations in excess mortality in ACCORD from intensive therapy (8,9), despite examining factors including hypoglycemia and hypoglycemia unawareness (which was actually less common among those who died in the intensive therapy arm) (10,11), age (12), cardiac autonomic dysfunction (13), weight gain (9), and rate of HbA_{1c} reduction (14). Although several factors in combination are thought to account for mortality HTEs, univariable subgroup analyses are not capable of identifying them and are subject to false-positive findings due to multiple testing (15,16).

Recently, the advancement of machine learning methods—particularly the approach of gradient forest analysis (17)—has aided the search for HTEs (Fig. 1). Gradient forest analysis can partition a trial population into subgroups characterized by multiple simultaneous characteristics (multivariable rather than univariable analysis), using cross-validation to reduce the likelihood of false-positive results (17). The gradient forest approach also inherently accounts for interactions among multiple variables (e.g., between age and HbA_{1c}) and is unbiased in predicting the difference in treatment effect between study arms, unlike older machine learning methods that can be biased and focus on the absolute rate of events (e.g., risk of mortality) rather than HTEs (e.g., how individual features affect the treatment's ability to reduce the risk of mortality) (17).

The objective of this study was to apply gradient forest analysis to identify subgroups of ACCORD participants with decreased or increased risk of all-cause mortality attributable to intensive therapy.
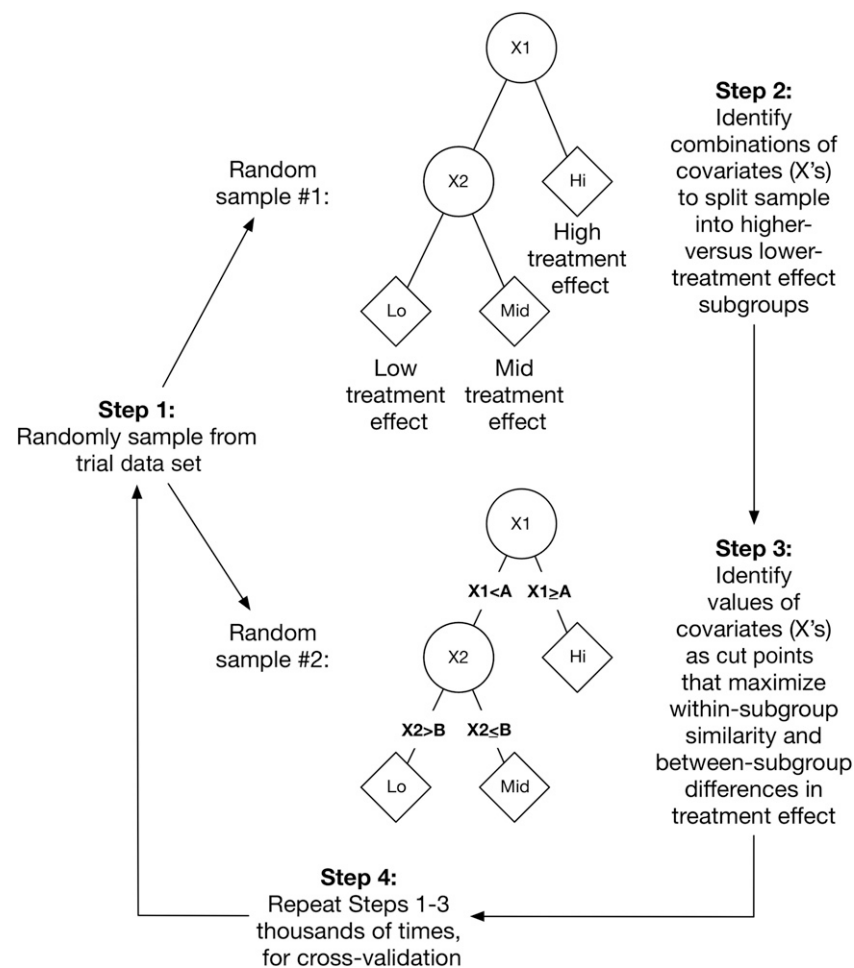
Figure 1—Conceptualization of gradient forest analysis to detect HTEs from trial data. Our implementation of gradient forest analysis involved repeated random sampling from both arms of the trial data set to compute the treatment effect—the difference in the probability of the primary outcome between the intensive and standard glycemic therapy arms—among subgroups of trial participants. After selecting subsamples of the trial data, our approach selected combinations of explanatory variables (X1, X2) from one subsection of data to divide the study population subsets with lower vs. higher treatment effects—in this case, all-cause mortality—when comparing intensive vs. standard therapy. We then used another subsection of data to update the preliminary values of the explanatory variables used to subdivide the population into final values that maximized between-group differences and minimized within-group differences in treatment effects among each subgroup. By using multiple subsections of data for the estimation of subgroups, the method produces unbiased estimates of HTE that are robust to outliers (17). The overall process is then repeated thousands of times to identify which variables and cut point values define consistent subgroups across thousands of random samplings from the trial data. The final subgroups chosen at the end of the decision tree are referred to as "leaves" of the tree.

RESEARCH DESIGN AND METHODS

Source of Data

ACCORD was a randomized, controlled trial of intensive versus standard glycemic control (open-label target of HbA$_{1c}$ <6.0% [42 mmol/mol] vs. 7.0–7.9% [53–63 mmol/mol], respectively), with a multifactorial design in which participants were additionally randomized to intensive versus standard lipid treatment (double-blinded assignment to fibrate plus statin or placebo plus statin, respectively), or intensive versus standard blood pressure treatment (open-label target of systolic blood pressure <120 mmHg or <140 mmHg, respectively) (2). The trial was conducted at 77 clinical sites in North America between January 2001 and June 2009. Participants in both arms received glucose-lowering medications. The glycemic control component of the trial was terminated early due to higher mortality in the intensive therapy arm, with a median on-protocol follow-up time of 3.7 years and a median on- plus off-protocol follow-up time of 4.9 years. The full duration of available data were used in this project. This analysis was approved by the Stanford University Institutional Review Board (e-Protocol #39321).

Participants

Participants (Supplementary Table 1) were 40–79 years old with type 2 diabetes, HbA$_{1c}$ ≥7.5% (58 mmol/mol), and prior evidence of cardiovascular disease (CVD) or risk factors for CVD (e.g., dyslipidemia, hypertension, smoking, or obesity; those without a prior cardiovascular event were between the ages of 55 and 79) (2,18,19). Exclusion criteria for ACCORD included BMI >45 kg/m^2, serum creatinine >1.5 mg/dL, or serious illnesses that might limit trial participation or life expectancy. Data from all study arms were included, with variables identifying glycemic, blood pressure, and lipid study arm to control for randomized therapy selection (15).

Outcome

The primary outcome for the current study was the difference in all-cause mortality between therapy arms, assessed from the point of enrollment to the time of study termination in June 2009. Mortality assessment in ACCORD was masked to therapy arm. The secondary outcome was the difference in composite microvascular events (including nephropathy, retinopathy, and neuropathy) between study arms, defined in ACCORD as renal failure, end-stage renal disease (dialysis), serum creatinine >3.3 mg/dL, photocoagulation or vitrectomy, or Michigan Neuropathy Screening Instrument score >2.0. As with mortality, assessment of microvascular events in ACCORD was masked to therapy arm. The secondary outcome was chosen to determine whether subgroups of participants identified based on HTEs for mortality exhibited similar HTEs in diabetes-related microvascular complications because the strongest support for intensive therapy has come from studies of reduced microvascular events. To help find groups with high mortality risk and low microvascular benefit, and vice versa, the decision tree based on the primary outcome was tested on the secondary outcome to determine whether the same features that predicted a higher or lower effect of intensive treatment on mortality would also predict a higher or lower effect of intensive treatment on microvascular events.

Predictors

Potential predictor variables for HTEs (itemized in Supplementary Table 1) included the subset of characteristics previously hypothesized to be related to cardiovascular or all-cause mortality among persons with type 2 diabetes: demographics (age, sex, race/ethnicity), study arm, type and number of glucose-lowering medications (including insulin use and oral glucose-lowering medication by class, individually, and in combination), diabetes history (years since diabetes diagnosis, hypoglycemia in prior 7 days), prior ulcer or amputation, history of eye disease or surgery, loss of vibratory sensation or monofilament sensation, biomarkers (HbA$_{1c}$, fasting blood glucose, hemoglobin glycosylation index [HGI] [20] [defined as observed − predicted HbA$_{1c}$ [%], where predicted HbA$_{1c}$ = 0.009 × fasting plasma glucose [mg/dL] + 6.8, using the single baseline fasting plasma glucose], lipid profile, serum creatinine, estimated glomerular filtration rate by the Modification of Diet in Renal Disease (MDRD) Study equation, serum potassium, urine microalbumin, urine creatinine, alanine aminotransferase, creatinine phosphokinase, systolic and diastolic blood pressure, heart rate, and BMI), and CVD covariates (tobacco smoking; atrial fibrillation or other arrhythmia by electrocardiogram; left ventricular hypertrophy by electrocardiogram; prior myocardial infarction, stroke, angina, bypass surgery, percutaneous coronary intervention, or other vascular procedure; and blood pressure medications, cholesterol medications, and anticoagulant/antiplatelet medications). HGI was included among the covariates because it was previously suggested as a potentially useful indicator of diabetes severity as well as a predictor of HTEs in mortality among persons with type 2 diabetes (20–22). Treatment arm (intensive vs. standard) is inherently part of the gradient forest analysis because the outcome is defined as difference in mortality between the two arms.

All predictor variables were taken from the baseline (prerandomization) study visit because our goal was to identify factors clinicians could use before the decision to set more, or less, intensive glycemic targets. Therefore, time-varying covariates were not incorporated into the analysis.

Sample Size

A total of 10,251 participants were included from the ACCORD trial, which includes the complete sample of participants enrolled.

Missing Data

Missing data were not imputed because <1% of data for any predictor variable were missing from the trial data set.

Statistical Analysis Method

To ensure transparency and reproducibility of the analysis, statistical code is linked at https://sdr.stanford.edu concurrent with publication. Our implementation of gradient forest analysis proceeded in four steps (Fig. 1). First, ACCORD trial data were divided in half randomly, with an equal number of intensive and standard glycemic control arm participants in each of the two data subsets. Second, variables were chosen by randomly sampling subsets of potential predictors to construct a decision tree made of those predictors that could split the first of the two subsamples of data into subgroups with higher and lower treatment effect (see Fig. 1). Treatment effect was defined as the absolute difference in the all-cause mortality rate between the intensive and standard therapy arms. Subgroups were required to be >5% of the overall study sample; we tested the consistency of the approach to ensure the same result

if we used limits of >1% to >8%. Third, once the initial decision tree was constructed from the first subsample of data, the values of each predictor that would define branches in the decision tree were refined using the second subsample of data so that the final subgroups at the bottom of the tree ("leaves" of the tree) had maximum between-group differences and minimum within-group differences in treatment effect. Refinement in the second data subset reduces the influence of outliers and helps produce unbiased HTE estimates (17). The overall approach was repeated 4,000 times from the first step to produce a "forest" of trees by repeated random resampling of the data (cross-validation). No change in estimated variable importance was observed beyond 4,000 trees. Variable importance was defined as the frequency with which a given variable was incorporated into a tree at the first, second, and further split points (i.e., a variable can change positions between trees, but variable selection for each position is tracked to monitor its importance). After the forest was constructed and cross-validated, the summary (average) decision tree was selected that separated participants into the subgroups that were most consistent across all trees in the forest (23).

To assess performance of the summary decision tree, the absolute risk difference in mortality was calculated between the intensive and standard glycemic control arms within each subgroup (leaf) of the trial population and compared across the subgroups (Q test for heterogeneity among subgroups and stratified log-rank test for trend in Kaplan-Meier all-cause mortality rates across subgroups). Absolute risk difference is the guideline-recommended outcome variable because it provides a clinically meaningful absolute, as opposed to relative, measure of effect (24–26). In addition, we estimated the Cox proportional hazards model for the outcome of mortality by treatment arm within each leaf, the hazard ratio of treatment, and the C statistic (area under the receiver operating characteristic curve) for discrimination of higher from lower overall mortality by leaf.

HTE models should not be confused for risk models (e.g., Cox models of the risk of mortality). An HTE model seeks to determine characteristics that are associated with treatment effectiveness. Hence, it models the difference in event rates

between treatment arms (the treatment effect) and tries to find the covariates that are associated with the treatment being more effective or less effective. A risk model, by contrast, finds correlates associated with a given outcome, such as identifying characteristics associated with the risk for mortality. Hence, it models the absolute event rate and tries to find the covariates (e.g., such as sex, blood pressure, etc.) that make overall mortality higher or lower; treatment may or may not be a covariate. A standard risk model does not specifically look for those factors that modify the treatment effect (i.e., interaction terms between study arm and covariates), whereas our gradient forest approach focuses exclusively on finding influential interaction terms, indicating those factors that modify the treatment effect. Furthermore, selection of an interaction term between treatment and effect modifiers may be reduced in significance by the larger effect on model fit and C statistic by the noninteracted terms and reveal only modification on a relative scale versus the absolute scale of the gradient forest approach.

Sensitivity Analyses
In sensitivity analyses, the summary decision tree was tested with the alternative outcome of difference in CVD mortality between study arms, defined in ACCORD as mortality suspected to be attributable to myocardial infarction, other acute coronary event, cardiovascular procedure, congestive heart failure, arrhythmia, or stroke. The effect of intensive therapy was notably larger (more adverse) for CVD mortality than for all-cause mortality in the ACCORD trial (2).

Analyses were performed in R 3.3.3 software (The R Project for Statistical Computing, Vienna, Austria).

RESULTS
Participants
Of the 10,251 study participants included in the analysis, 718 died during study follow-up from all causes, including 327 participants (6.4%) in the standard therapy arm and 391 participants (7.6%)in the intensive therapy arm. CVD was attributed as the cause of death for 331 participants (3.2% of participants, 46.1% of deaths), including 144 (2.8% of participants) in the standard glycemic therapy arm and 187 (3.6% of participants) in the intensive glycemic therapy arm. As in the original ACCORD

publication (2), the hazard ratio of treatment was 1.17 (95% CI 0.98, 1.40) for all-cause mortality in the intensive versus standard glycemic group overall, after including all predictor covariates in a standard Cox regression model, and 1.20 (95% CI 1.04, 1.39) without predictor covariates included.

Model Specification
The summary decision tree (Fig. 2) separated the ACCORD population by variation in all-cause mortality rate differences between the standard and intensive therapy arms. The first split of the tree was defined by the HGI, which was selected as the key splitting variable in 2,390 of 4,000 trees (59.8%). For participants with low HGI (<0.44, or 75% of the study sample), the next split was defined by BMI, which was selected as a subsequent splitting variable in 2,322 of 4,000 trees (58.1%). The group with a low BMI (<30 kg/m^2, a derived value rounded to the nearest kg/m^2) was further split by age (<61 years), which was selected in 1,814 of the 4,000 trees (45.4%). The three variables defining the decision tree were available for 9,801 of the 10,251 ACCORD trial participants (95.6%).

Model Performance
The summary decision tree split the study sample into groups with significantly different risk for all-cause mortality from intensive glycemic therapy, as reported in Table 1 (P < 0.001 by the Q test for heterogeneity in absolute mortality risk difference between intensive vs. standard therapy among the four groups, and P < 0.001 by the stratified log-rank test for a trend in absolute mortality difference from subgroup 1 through subgroup 4).

Subgroup (leaf) 1 had 877 participants (8.6% of the 10,251-participant total sample) and was defined by HGI <0.44, BMI <30 kg/m^2, and age <61 years old. Subgroup 1 had an absolute mortality rate reduction (benefit) of 2.3% from intensive glycemic therapy (95% CI 0.2 to 4.5 decrease; hazard ratio 0.41; 95% CI 0.17, 0.98; P = 0.038 by the log-rank test adjusting for censoring). Participants in subgroup 1 had a number needed to treat (NNT) of 43 over 5 years to observe 1 less death with intensive rather than standard glycemic therapy.

Subgroup (leaf) 2 had 1,717 participants (16.7% of sample) and was defined
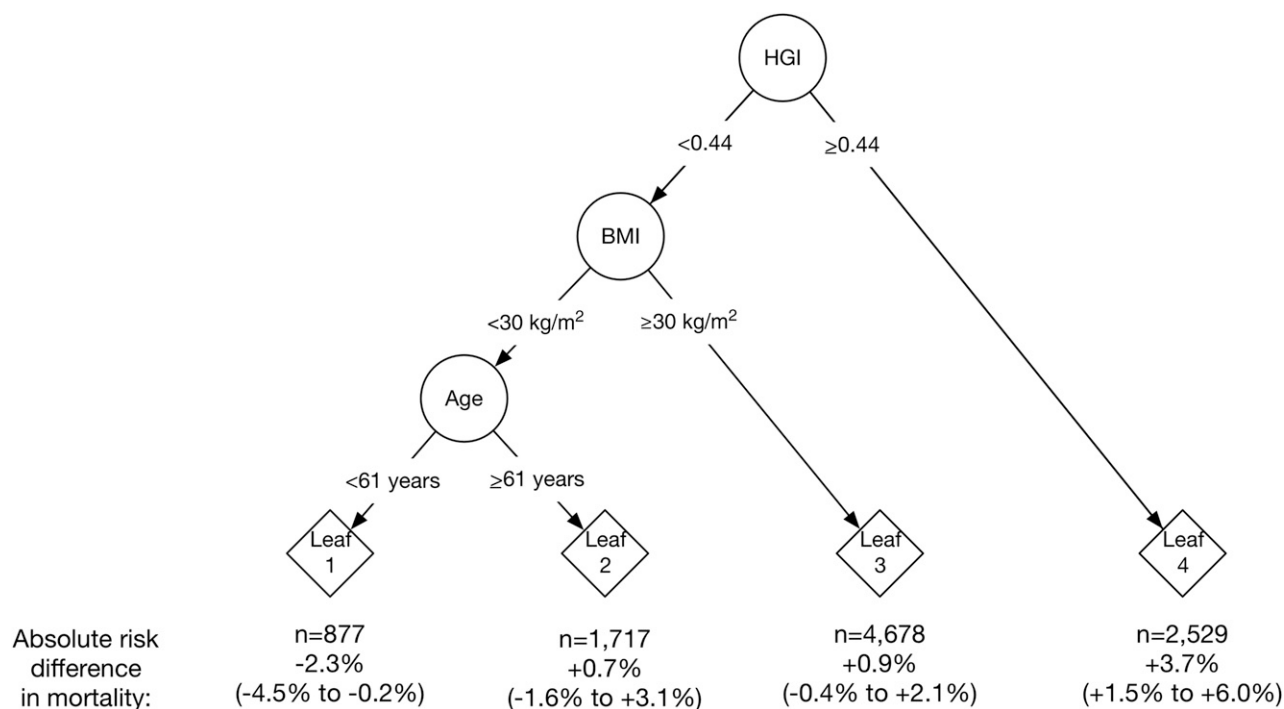
Figure 2—Summary risk stratification decision tree developed to identify the absolute change in risk of all-cause mortality among persons with type 2 diabetes subject to intensive therapy, based on baseline characteristics of individual participants in the ACCORD trial (2001–2009, N = 10,251). Negative values indicate reduced absolute mortality (benefit from intensive glycemic control), whereas positive values indicate increased absolute mortality (harm from intensive glycemic control).

by HGI <0.44, BMI <30 kg/m^2, and age ≥61 years old. Subgroup 2 had no significant absolute mortality rate reduction or increase, with an absolute risk increase of 0.7% from intensive glycemic therapy (95% CI 1.6 decrease to 3.1 increase; hazard ratio 1.11, 95% CI 0.77, 1.60; P = 0.560).

Subgroup (leaf) 3 had 4,678 participants (45.6% of sample) and was defined by HGI <0.44 and BMI ≥30 kg/m^2. Subgroup 3 had no significant absolute mortality rate reduction or increase, with an absolute risk increase of 0.9% from intensive glycemic therapy (95% CI 0.4 decrease to 2.1 increase) and a hazard ratio of 1.12 (95% CI 0.91, 1.50; P = 0.220).

Subgroup (leaf) 4 had 2,529 participants (24.7% of sample) and was defined by HGI ≥0.44. Subgroup 4 had an absolute mortality rate increase of 3.7% from intensive glycemic therapy (95% CI 1.5 to 6.0 increase) and a hazard ratio of 1.57 (95% CI 1.20, 2.04; P < 0.001). Participants in subgroup 4 had a number needed to harm of 27 over 5 years associated with 1 additional death in the intensive than standard glycemic therapy arm.

Figure 3 illustrates the survival curves among the intensive and standard glycemic therapy arms of ACCORD, stratified by the subgroups. Supplementary Table 2 lists the other clinical features among the subgroups by arm, revealing that covariates were balanced across the therapy arms within each subgroup. Hence, imbalance in important covariates between arms did not result from the stratification into subgroups, meaning that the gradient forest analysis did not produce confounding by measured covariates. Critically, Supplementary Table 2 also reveals that because no single predictor variable could explain the subgroups, the decision tree did not capture features that would be otherwise obvious from a univariable subgroup analysis; rather, the multivariate machine learning analysis had the power to reveal variations in mortality that would not be detectable to univariable subgroup analyses along any of the measured variables in the study. Overall, the out-of-the-bag error rate of the model, a measure of the prediction error during out-of-sample cross-validation, was low, with a value of 5.6%.

We evaluated whether the secondary outcome of composite microvascular events varied among the subgroups

(Supplementary Table 3 and Supplementary Fig. 1). The average decrease in microvascular outcomes was nonsignificant for all four subgroups, consistent with the overall results of the ACCORD trial (24). However, the average outcomes were better for subgroup 1 (absolute risk decrease of 4.2%, 95% CI 10.6 decrease to 2.1 increase, P = 0.15) than for subgroup 4 (absolute risk decrease of 2.3%, 95% CI 6.1 decrease to 1.5 increase, P = 0.60).

In sensitivity analyses, we evaluated the summary decision tree with the outcome of CVD mortality (Supplementary Table 4 and Supplementary Fig. 2). As with absolute risk differences in all-cause mortality, absolute risk differences in CVD mortality between intensive and standard glycemic therapy differed significantly between the four subgroups (P < 0.001 for heterogeneity and for trend). Subgroup 1 had an absolute cardiovascular mortality risk decrease of 1.7% in the intensive therapy arm (95% CI 0.2 to 3.2 decrease, P = 0.027), and subgroup 4 had an absolute cardiovascular mortality risk increase of 2.3% in the intensive therapy arm (95% CI 0.6 to 3.9 increase, P = 0.004).

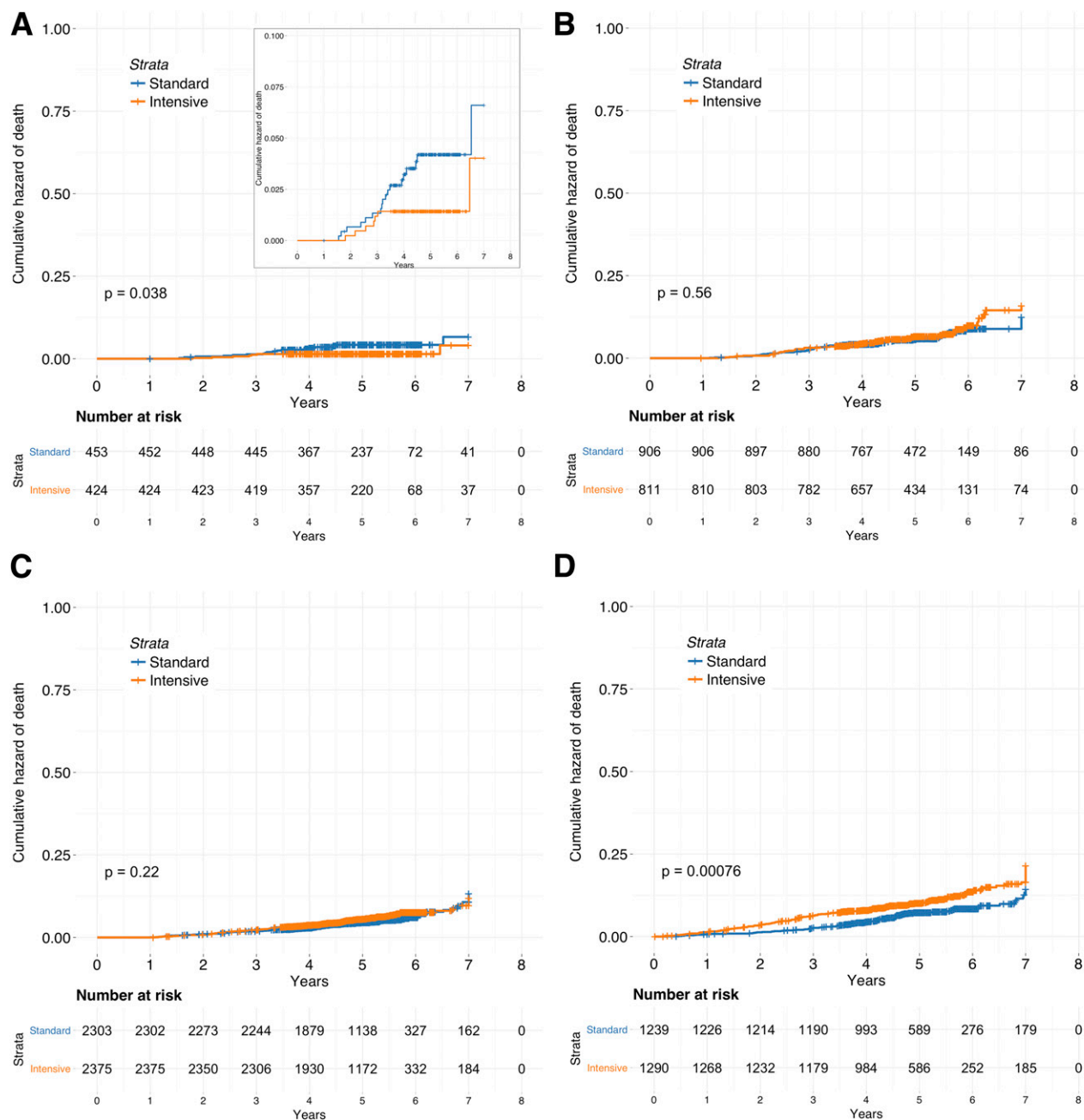**Table 1—Change in absolute risk of all-cause mortality from intensive versus standard glycemic control, among subgroups identified by gradient forest analysis of the ACCORD trial (2001–2009, N = 10,251)**

Group	Total deaths (N = 9,801 of 10,251 with variables to stratify risk), n (%)		Total deaths (N = 9,801 of 10,251 with variables to stratify risk), n (%)	Deaths among intensive therapy (N = 4,900 of 5,128 with variables to stratify risk), n (%)	Deaths among standard therapy (N = 4,901 of 5,123 with variables to stratify risk), n (%)	Intensive vs. standard treatment, hazard ratio (95% CI), C statistic (95% CI)	Absolute risk difference, % (95% CI)	P value (log-rank test) of difference in event rates between arms, within leaf	Stratified log-rank test (difference in treatment effect across leaves)
	Intensive therapy	Standard therapy							
Leaf 1 (leftmost in Fig. 2) (HGI <0.44, BMI <30 kg/m², age <61 years)	424	453	25 (2.9)	7 (1.7)	18 (4.0)	0.41 (0.17, 0.98), 0.64 (0.52, 0.76)	−2.3 (−4.5 to −0.2)	0.04	
Leaf 2 (HGI <0.44, BMI <30 kg/m², age ≥61 years)	811	906	116 (6.8)	58 (7.2)	58 (6.4)	1.11 (0.77, 1.60), 0.62 (0.56, 0.67)	0.7 (−1.6 to 3.1)	0.56	
Leaf 3 (HGI <0.44, BMI ≥30 kg/m²)	2,375	2,303	250 (5.3)	137 (5.8)	113 (4.9)	1.12 (0.91, 1.50), 0.64 (0.60, 0.68)	0.9 (−0.4 to 2.1)	0.22	<0.001
Leaf 4 (rightmost in Fig. 2) (HGI ≥0.44)	1,290	1,239	234 (9.3)	143 (11.1)	91 (7.3)	1.57 (1.20, 2.04), 0.66 (0.62, 0.71)	3.7 (1.5 to 6.0)	<0.001	

See Fig. 2 for visualization of subgroups. Note that the hazard ratio of intensive vs. standard treatment and the C statistic (area under the receiver operating characteristic curve) for discrimination of higher from lower overall mortality by leaf was estimated the Cox proportional hazards model for the outcome of mortality by treatment within each leaf.

CONCLUSIONS

We sought to inform clinical decisions regarding the safety of intensive glycemic therapy among patients with type 2 diabetes and elevated CVD risk by identifying HTE in all-cause mortality within the ACCORD trial. We found that by using the covariates of HGI, age, and BMI, we could classify participants in the ACCORD trial into subgroups with clinically meaningful differences in mortality attributable to intensive glycemic therapy. The mean all-cause mortality rate among individuals with diabetes in the U.S. is ~6% over 5 years, so an absolute risk increase of 4% or an absolute risk reduction of 2% is clinically meaningful (25). Approximately 25% (n = 2,529) of participants belonged to a subgroup experiencing increased mortality attributable to intensive glycemic therapy, whereas 9% (n = 877) belonged to a subgroup that experienced reduced mortality attributable to intensive glycemic therapy. We did not find that hypoglycemia, medication classes, number of medications, combinations of medications, baseline diabetes complications, or cardiovascular risk factors could explain the HTEs from intensive glycemic therapy. We also did not find a trade-off between microvascular and mortality risk, because the patients with the highest mortality risk from intensive therapy also had the least evidence of microvascular benefit, and vice versa.

Our findings support and extend prior studies of glycemic control in diabetes management. We found that despite the average treatment effect of higher mortality, there were some groups that may have benefited from, along with some that were likely harmed by, intensive glycemic therapy; nearly two-thirds (n = 6,395) experienced neither benefit nor harm. Because the risk of benefit and of harm varies among individuals with type 2 diabetes, our results support current guidelines that advocate for individualized treatment decisions and also help such guidelines to be made operational in clinical practice (1). Clinically, the decision tree we developed through a data-driven multivariate subgroup analysis uses readily available clinical data and may assist clinician-patient discussions about glycemic therapy. Although the ACCORD HbA$_{1c}$ target of <6.0% (42 mmol/mol) is not guideline recommended, many patients are currently treated to <6.5%

Figure 3—Survival curves for all-cause mortality among subsets identified by each subgroup in the decision tree (see Fig. 2 for subgroups). P values are from the stratified log-rank test adjusted for censorship of Kaplan-Meier all-cause mortality rates among the intensive vs. standard glycemic therapy arm. A: Leaf 1 (leftmost in Fig. 2) (HGI <0.44, BMI <30 kg/m², and age <61 years). B: Leaf 2 (HGI <0.44, BMI <30 kg/m², and age ≥61 years). C: Leaf 3 (HGI <0.44 and BMI ≥30 kg/m²). D: Leaf 4 (rightmost in Fig. 2) (HGI ≥0.44).

(48 mmol/mol, the achieved mean in the intensive therapy arm) with regimens other than metformin alone (5,19). Because ~25% of ACCORD-eligible patients were observed to have high risk of harm from intensive therapy, deescalation of glycemic therapy may be warranted for some patients. Our study also adds to a growing body of literature, including a prior study using ACCORD data, that a high HGI may be an important indicator of diabetes severity as well as a predictor of

HTEs in mortality among persons with type 2 diabetes (20–22). A higher HGI may indicate higher postprandial glucose levels and increased glycemic variability. Notably, the HGI in this report does not require that mean glucose levels be determined by continuous glucose monitoring, as is common in studies of type 1 diabetes. Rather HGI was calculated using a single HbA$_{1c}$ and fasting plasma glucose measurement, offering potential convenience for clinical use.

More broadly, these results point toward the application of innovative methods for the detection of HTEs from clinical trial data. Our findings highlight the point that trial summary statistics, which are averages, may obscure clinically important heterogeneities and that the rigorous application of machine learning methods with conservative cross-validation approaches may aid in finding consistent subgroups that experience substantial differences in treatment effects.

Extensive theoretical and empirical research suggests that the ability of conventional univariable subgroup analyses to detect clinically important heterogeneity in treatment effects is very limited (26–28). Previous studies of HTEs in ACCORD data have considered single variables, finding that hypoglycemia and cardiac autonomic dysfunction did not explain the harm of intensive therapy (10,11,13). The machine learning method accounting for multiple simultaneous covariates and interactions between them was therefore able to explain the variation in mortality better than previous univariable analyses. We in fact found in our sensitivity analyses that no single covariate would be able to distinguish the subgroups, and therefore, our multivariable machine learning analysis had the power to explain variations that were not possible to find with traditional univariable subgroup analyses.

A prior analysis of age as a source of HTEs found that younger age was associated with increased harm (12). Our finding that younger age, in combination with lower BMI and HGI, is instead associated with benefit may represent the interaction of factors not considered in univariable analyses. In general, considering several factors in combination may be required to explain clinically important variations in benefit and harm seen in clinical trials. Consequently, multivariate HTE modeling has been increasingly recommended (15,16,29). Our data-driven approach also adjusts for type I error due to multiple hypothesis testing, a major disadvantage of traditional subgroup analysis methods. We used rigorous cross-validation to reduce the chance of false-positive findings.

Our analysis nevertheless has important limitations. As a result of the ACCORD trial being stopped early, we could assess only shorter-term outcomes. Further, the ACCORD trial was conducted before the widespread availability of sodium–glucose cotransporter 2 and glucagon-like peptide 1 agents, which have cardiovascular benefits that affect the risk of mortality with glycemic therapy (30–33). In addition, because we wanted a clinical decision tree that was useful in practice, we focused on pretreatment characteristics rather than time-varying covariates, which may be more useful in predicting outcomes over time but are also more complex for

clinicians to use. Next, although we used methods to minimize the risk of type I error and did not observe imbalance in covariates within subgroups, our study is nevertheless a post hoc analysis of a single trial. With machine learning methods, as with correlative statistical methods in general, variable selection does not prove causality, and the variables selected may only be surrogates for more complex physiological processes. HGI is a summary measure that may not have a definitive physiological meaning and can be calculated in alternative ways; here, it serves as a useful and readily calculable marker of complex physiological processes and was found to separate the variation in mortality better than alternative covariates. HGI thus likely reflects a complex underlying heterogeneity in treatment effect. Explaining mechanistically the physiological relationships that underlie the HTEs observed is not possible from the available data, although they are broadly consistent with clinical observation and point to areas for further study (34). In addition, the number of deaths among the standard and intensive therapies in leaf 1 were too small (7 and 18 subjects). Finally, it is important to note that these results naturally apply to the population that met inclusion criteria for ACCORD, which includes people with type 2 diabetes with HbA$_{1c}$ of $\geq$7.5%, who were between the ages of 40 and 79 years and had CVD or were between the ages of 55 and 79 years and had anatomical evidence of significant atherosclerosis, albuminuria, left ventricular hypertrophy, or at least two additional risk factors for CVD, such as dyslipidemia, hypertension, tobacco smoking, or obesity.

Our study suggests several directions for future work. Because only internal validation was done in this report, prospectively validating the decision tree on an independent trial data set and on population-based observational data would help assess the generalizability of our findings. Ultimately, it will be important to evaluate the effect of using the decision tree on clinical practice and patient outcomes. More generally, HTEs are likely to be the norm, rather than the exception, in many areas of investigation. Therefore, it may be advantageous to design trials that can identify HTEs up front, rather than relying on post hoc analysis as we have done here. A prior simulation study revealed that alternative trial

designs, which randomize persons in a stepwise fashion to incrementally higher levels of therapy intensification, could increase statistical power to detect HTEs and provide more granular estimates of treatment benefit or harm (28). Finally, the analysis suggests that HGI may be a useful clinical indicator of risk and advanced diabetes, necessitating future prospective study as a useful clinical biomarker.

Clinicians may use HGI, age, and BMI to help individualize decisions about glycemic control among people with type 2 diabetes. This may lead to deescalation of therapy for many patients while also identifying patients who do not face increased all-cause mortality risk from their current glycemic therapy. Further, the methods used in this study offer a principled way to help inform individualized care using data from randomized trials. The application of similar methods may enable us to learn more from the contribution that clinical trial participants make, bringing us closer to the goal of personalized medicine.

References

1. American Diabetes Association. Glycemic targets. Sec. 6. In *Standards of Medical Care in Diabetes—2017*. Diabetes Care 2017;40(Suppl. 1):S48–S56

2. Action to Control Cardiovascular Risk in Diabetes Study Group; Gerstein HC, Miller ME, Byington RP, et al. Effects of intensive glucose lowering in type 2 diabetes. N Engl J Med 2008;358:2545–2559

3. Zoungas S, Arima H, Gerstein HC, et al.; Collaborators on Trials of Lowering Glucose (CONTROL) group. Effects of intensive glucose control on microvascular outcomes in patients with type 2 diabetes: a meta-analysis of individual participant data from randomised controlled trials. Lancet Diabetes Endocrinol 2017;5:431–437

4. Sussman JB, Kerr EA, Saini SD, et al. Rates of deintensification of blood pressure and glycemic medication treatment based on levels of control and life expectancy in older patients with diabetes mellitus. JAMA Intern Med 2015;175: 1942–1949

5. Lipska KJ, Ross JS, Miao Y, Shah ND, Lee SJ, Steinman MA. Potential overtreatment of diabetes mellitus in older adults with tight glycemic control. JAMA Intern Med 2015;175:356–362

6. McCoy RG, Van Houten HK, Ross JS, Montori VM, Shah ND. HbA1c overtesting and overtreatment among US adults with controlled type 2 diabetes, 2001-13: observational population based study. BMJ 2015;351:h6138.

7. 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; 359:1577–1589

8. Skyler JS, Bergenstal R, Bonow RO, et al.; American Diabetes Association; American College of Cardiology Foundation; American Heart Association. Intensive glycemic control and the prevention of cardiovascular events: implications of the ACCORD, ADVANCE, and VA diabetes trials: a position statement of the American Diabetes Association and a scientific statement of the American College of Cardiology Foundation and the American Heart Association. Diabetes Care 2009;32:187–192

9. Riddle MC. Counterpoint: intensive glucose control and mortality in ACCORD–still looking for clues. Diabetes Care 2010;33:2722–2724

10. Bonds DE, Miller ME, Bergenstal RM, et al. The association between symptomatic, severe hypoglycaemia and mortality in type 2 diabetes: retrospective epidemiological analysis of the ACCORD study. BMJ 2010;340:b4909

11. Seaquist ER, Miller ME, Bonds DE, et al.; ACCORD Investigators. The impact of frequent and unrecognized hypoglycemia on mortality in the ACCORD study. Diabetes Care 2012;35:409–414

12. Miller ME, Williamson JD, Gerstein HC, et al.; ACCORD Investigators. Effects of randomization to intensive glucose control on adverse events, cardiovascular disease, and mortality in older versus younger adults in the ACCORD trial. Diabetes Care 2014;37:634–643

13. Pop-Busui R, Evans GW, Gerstein HC, et al.; Action to Control Cardiovascular Risk in Diabetes Study Group. Effects of cardiac autonomic dysfunction on mortality risk in the Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial. Diabetes Care 2010;33:1578–1584

14. Riddle MC, Ambrosius WT, Brillon DJ, et al.; Action to Control Cardiovascular Risk in Diabetes Investigators. Epidemiologic relationships between A1C and all-cause mortality during a median 3.4-year follow-up of glycemic treatment in the ACCORD trial. Diabetes Care 2010;33:983–990

15. 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: 163–169.

16. 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

17. Athey S, Imbens G. Recursive partitioning for heterogeneous causal effects. Proc Natl Acad Sci U S A 2016;113:7353–7360

18. ACCORD Study Group; Buse JB, Bigger JT, Byington RP et al. Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial: design and methods. Am J Cardiol 2007;99:21i–33i

19. Gerstein HC, Riddle MC, Kendall DM, et al.; ACCORD Study Group. Glycemia treatment strategies in the Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial. Am J Cardiol 2007;99: 34i–43i

20. Hempe JM, Liu S, Myers L, McCarter RJ, Buse JB, Fonseca V. The hemoglobin glycation index identifies subpopulations with harms or benefits from intensive treatment in the ACCORD trial. Diabetes Care 2015;38:1067–1074

21. van Steen SC, Schrieks IC, Hoekstra JB, et al.; AleCardio study group. The haemoglobin glycation index as predictor of diabetes-related complications in the AleCardio trial. Eur J Prev Cardiol 2017;24:858–866

22. McCarter RJ, Hempe JM, Gomez R, Chalew SA. Biological variation in HbA1c predicts risk of retinopathy and nephropathy in type 1 diabetes. Diabetes Care 2004;27:1259–1264

23. Banerjee M, Ding Y, Noone AM. Identifying representative trees from ensembles. Stat Med 2012;31:1601–1616

24. Ismail-Beigi F, Craven T, Banerji M, et al. Effect of intensive treatment of hyperglycaemia on microvascular complications of type 2 diabetes in ACCORD: a randomised trial. Lancet 2010;376: 419–430

25. National Center for Health Statistics. *Health, United States, 2016: With Chartbook on Long-Term Trends in Health.* Washington, D.C., Department of Health and Human Services, 2017

26. VanderWeele TJ, Knol MJ. Interpretation of subgroup analyses in randomized trials: heterogeneity versus secondary interventions. Ann Intern Med 2011;154:680–683

27. Wallach JD, Sullivan PG, Trepanowski JF, Sainani KL, Steyerberg EW, Ioannidis JP. Evaluation of evidence of statistical support and corroboration of subgroup claims in randomized clinical trials. JAMA Intern Med 2017; 177:554–560

28. 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;166:354–360

29. 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

30. Marso SP, Bain SC, Consoli A, et al.; SUSTAIN-6 Investigators. Semaglutide and cardiovascular outcomes in patients with type 2 diabetes. N Engl J Med 2016;375:1834–1844

31. Marso SP, Daniels GH, Brown-Frandsen K, et al.; LEADER Steering Committee; LEADER Trial Investigators. Liraglutide and cardiovascular outcomes in type 2 diabetes. N Engl J Med 2016;375: 311–322

32. Zinman B, Wanner C, Lachin JM, et al.; EMPA-REG OUTCOME Investigators. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. N Engl J Med 2015;373:2117–2228

33. Neal B, Perkovic V, Mahaffey KW, et al.; CANVAS Program Collaborative Group. Canagliflozin and cardiovascular and renal events in type 2 diabetes. N Engl J Med 2017;377:644–657

34. American Diabetes Association. Pharmacologic approaches to glycemic treatment. Sec. 8. In *Standards of Medical Care in Diabetes—2017.* Diabetes Care 2017;40(Suppl. 1):S64–S74