Point: Intensive Glycemic Control and Mortality in ACCORD—A Chance Finding?

The Action to Control Cardiovascular Risk in Diabetes (ACCORD) study was designed to assess whether intensive glucose management aimed at a normal A1C (<6%) versus standard therapy aimed at an A1C of 7–8% would reduce the risk of cardiovascular disease as reflected by the incidence of a major adverse cardiac event (MACE) such as cardiovascular death, nonfatal myocardial infarction, or nonfatal stroke (1). The study enrolled 10,251 patients, and a planned mean follow-up of 5.6 years was expected to provide 89% power to detect a 15% reduction in MACE risk with intensive versus conventional therapy. Patients received any one of six antihyperglycemic agents or a variety of insulins alone or in combination. Eligible subjects were also randomized to intensive versus conventional lipid and/or blood pressure control. However, after a mean of 3.6 years of follow-up, when 62.5% of the planned information had been accrued, the study was terminated by the Data and Safety Monitoring Board (DSMB) in the presence of an increased incidence of mortality with intensive therapy: 257 versus 203 deaths with standard therapy, corresponding to an adjusted hazard ratio (HR) of 1.22 (95% CI [1.01–1.46], $P = 0.04$). The database used to generate the final report to the DSMB was frozen and then used as the basis for analyses that were subsequently published (2).

The initial report (2) stated that "differences in the use of drugs (including rosiglitazone), weight change, and other factors did not identify an explanation for the mortality finding." Since then the authors have searched in earnest for a biological mechanism that would account for this increase in mortality. None has been found. Herein I present a synopsis of these findings and then suggest another possible mechanism—that the observed excess mortality could have been a chance finding.

Baseline factors

Subgroup analyses within strata defined by baseline factors can assess whether the baseline factor modifies the effect of treatment as reflected by a test of homogeneity of treatment effects over strata or a test of group by covariate interaction. As pointed out by Wang et al. (3), such assessments should adjust for the number of subgroup factors since whenever a sample is split in two, by chance the treatment group difference in one stratum will be greater than that in another. Such analyses in ACCORD (4) demonstrated nominally significant heterogeneity for 3 of the 38 factors examined. The Intensive: Standard (I:S) HR was 1.95 among those with a history of diabetic neuropathy by self-report compared with HR = 0 for those without, nominal $P = 0.0008$ between strata, and $P = 0.031$ when adjusted for 38 tests. This suggests that those with preexisting neuropathy may be a vulnerable subset but alone it does not identify a mechanism for the increased risk of mortality with intensive therapy. The I:S HR was greater among those using aspirin versus those not using aspirin (1.45 vs. 0.95, $P = 0.031$) and was greater among those with a baseline A1C >8.5% versus those with a baseline A1C ≤8.5% (1.6 vs. 0, $P = 0.044$). Neither remains close to being significant when adjusted for 38 such analyses.

Hypoglycemia

Intensive therapy increased the risk of severe hypoglycemia requiring assistance (HA) and requiring medical assistance (HMA) approximately threefold versus standard therapy (5). Contrary to expectation, the risk of hypoglycemia in both treatment groups increased as the A1C level increased. Further, within both treatment groups, those with at least one episode of HA had a greater risk of mortality than those with no HA, HR = 1.79 in intensive, 2.93 in standard, both highly significant.

In additional epidemiological analyses (6), among the 9,122 subjects who did not experience any hypoglycemia, there was a nonsignificant increased risk of mortality with intensive versus standard therapy, the I:S HR = 1.21 (0.99–1.48). Among the 1,072 who experienced HA, the I:S HR decreased as the number of such episodes increased, HR = 0.84 after one episode, 0.71 after two, and 0.44 after three. The test of homogeneity, however, was not statistically significant ($P = 0.23$). A similar pattern was observed for HMA for which the test of homogeneity was nominally significant at $P = 0.05$.

Glycemia (A1C)

Since intensive therapy was designed to lower glucose, as measured by A1C, an association with A1C was carefully explored in further epidemiological analyses (7). Summary measures of glycemia were employed as time-dependent covariates in regression models, as in like analyses of the Diabetes Control and Complications Trial (DCCT) (8) and the UK Prospective Diabetes Study (UKPDS) (9).

The decrease in A1C over the first 4 months and the decrease over the first year (1-year decrease) had no association with mortality risk (HR = 1.0 and 1.02 per A1C unit decrease, respectively). The authors also presented a spline-smoothed estimate of the mortality risk as a function of the 1-year A1C decrease that was flat over the range of values in both groups, except that there was the suggestion of an increased risk among intensive therapy subjects who were unable to reduce their A1C levels. However, the regression model estimated an HR = 0.86 per unit (1%) A1C decrease over the first year with intensive therapy versus 0.83 per unit decrease in the standard group. Thus, there is no suggestion that a greater decrease in A1C over the first year is associated with an increased risk.

The updated mean A1C (average A1C) at the time of each death (or the end of follow-up) was significantly associated with mortality in the combined cohort with a 20% increase in risk of mortality per unit increase in A1C (HR = 1.2, $P = 0.0002$), whereas the most recent (last) value had no association (HR = 1.06). A spline-smoothed estimate of the log(HR) per unit increase in the average A1C value in the standard group was essentially flat (no association) resulting in a model-estimated HR of 0.98 per unit increase in A1C. Conversely, in the intensive group, the log(HR) increased as a linear function

of the A1C over the range of A1C values, i.e., a positive rather than a negative association, resulting in an HR of 1.66 per unit A1C increase ($P < 0.0001$), the difference in the risk gradients between groups also being statistically significant ($P = 0.0007$).

Whether a given covariate represents the mechanism by which treatment has an effect on an outcome can be assessed by the relative magnitude of the treatment effect in regression models without and with adjustment for the covariate, say T_u and T_a, respectively. The metric could either be the regression coefficient or the test statistic value. The percentage of the treatment effect explained is then quantified by the percentage reduction in the treatment effect, i.e., $100 \times (T_u - T_a)/T_u$. An example was provided by the DCCT (10). However, if the treatment effect is unchanged or increases, then the covariate is not related to the mechanism of the treatment effect. After adjustment for the average A1C over time, the I:S HR = 1.82 increased markedly from the unadjusted HR of 1.22 in Gerstein and colleagues (2) and 1.25 in Riddle et al. (7), adjusted for other factors. This suggests that statistically the estimated excess risk of mortality with intensive versus standard therapy would be greater if the intensive and standard therapy groups had the same average A1C over time. Clearly the mean A1C is not related to the mechanism that led to the increased incidence of mortality.

Chance

Under Occam's razor or the principle of parsimony, chance is the simplest hypothesis that is consistent with the failure thus far to identify a plausible mechanism for the increased risk of mortality with intensive therapy. A chance finding or a false positive type I error is random noise that is neither predictable nor correlated with other variables such that it would be expected that no mechanism could be identified. Statistically, the finding of increased risk of mortality (HR = 1.22, $P = 0.04$) was the third most significant of 5 secondary outcomes assessed in Gerstein and colleagues (2), there being a 24% risk reduction of nonfatal myocardial infarction ($P = 0.004$) and a 35% risk increase of cardiovascular death ($P = 0.02$) with intensive versus standard therapy. Thus, an adjusted P value for mortality would be $0.04 \times 3 = 0.12$. Further, the DSMB met approximately every 6 months over a 6-year period. Using standard calculations for repeated significance tests (11) with 10 equally spaced looks and 62.5% information at the 10th look, the probability that a single outcome would have $P \leq 0.04$ at one of these looks is 0.16. Thus, the total type I error probability associated with the mortality difference could be as high as 0.48.

Further, Figure A2 in the online appendix for Riddle et al. (7) shows that the risk of mortality with intensive therapy was negligibly different between groups at years 1 and 2, substantially higher with intensive therapy in year 3, lower than that with standard therapy in years 4 and 5, and then greater again at year 6 with no difference at year 7. There is no systematically increased risk over time, consistent with chance as the mechanism.

However, this does not mean that the investigators do not need to search further for a mechanism. If indeed a mechanism can be identified, it could certainly save lives. The possibilities are limitless, and some, like the hypothesis of a "toxic soup" of multiple agents, may be difficult to assess; but the effort should continue.

DSMB

While the type I (false positive) error probability is a principal concern when considering early termination for effectiveness, and the type II (false negative) error probability for futility, no statistical boundary or stopping rule is generally employed for safety. The latter relies on the judgment of the DSMB as a whole based on the totality of evidence available. An excess of 54 deaths with intensive therapy, even though perhaps not statistically compelling, would cause grave concern. Thus, conjecture that this excess could have occurred by chance is not intended, nor should it be in any way construed, to imply that the DSMB acted unwisely when it terminated the trial. Their sole responsibility was to weigh the evidence for potential harm so as to protect the safety of the subjects enrolled in the trial. It was not to prove that intensive therapy was harmful.

JOHN M. LACHIN, SCD

From the Biostatistics Center, The George Washington University, Rockville, Maryland.

Corresponding author: John M. Lachin, jml@bsc.gwu.edu.

DOI: 10.2337/dc10-1506

Acknowledgments— No potential conflicts of interest relevant to this article were reported.

References

1. Action to Control Cardiovascular Risk in Diabetes Study Group. Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial: design and methods. Am J Cardiol 2007;99 (Suppl.):21i–33i
2. Action to Control Cardiovascular Risk in Diabetes Study Group, Gerstein HC, Miller ME, Byington RP, Goff DC Jr, Bigger JT, Buse JB, Cushman WC, Genuth S, Ismail-Beigi F, Grimm RH Jr, Probstfield JL, Simons-Morton DG, Friedewald WT. Effects of intensive glucose lowering in type 2 diabetes. N Engl J Med 2008;358:2545–2559
3. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine–reporting of subgroup analyses in clinical trials. N Engl J Med 2007;357:2189–2194
4. Calles-Escandón J, Lovato LC, Simons-Morton DG, Kendall DM, Pop-Busui R, Cohen RM, Bonds DE, Fonseca VA, Ismail-Beigi F, Banerji MA, Failor A, Hamilton B. Effect of intensive compared with standard glycemia treatment strategies on mortality by baseline subgroup characteristics: the Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial. Diabetes Care 2010;33:721–727
5. Miller ME, Bonds DE, Gerstein HC, Seaquist ER, Bergenstal RM, Calles-Escandon J, Childress RD, Craven TE, Cuddihy RM, Dailey G, Feinglos MN, Ismail-Beigi F, Largay JF, O'Connor PJ, Paul T, Savage PJ, Schubart UK, Sood A, Genuth S; ACCORD Investigators. The effects of baseline characteristics, glycaemia treatment approach, and glycated haemoglobin concentration on the risk of severe hypoglycaemia: post hoc epidemiological analysis of the ACCORD study. BMJ 2010;340:b5444
6. Bonds DE, Miller ME, Bergenstal RM, Buse JB, Byington RP, Cutler JA, Dudl RJ, Ismail-Beigi F, Kimel AR, Hoogwerf B, Horowitz KR, Savage PJ, Seaquist ER, Simmons DL, Sivitz WI, Speril-Hillen JM, Sweeney ME. The association between symptomatic, severe hypoglycaemia and mortality in type 2 diabetes: retrospective epidemiological analysis of the ACCORD study. BMJ 2010;340:b4909
7. Riddle MC, Ambrosius WT, Brillon DJ, Buse JB, Byington RP, Cohen RM, Goff DC Jr, Malozowski S, Margolis KL, Probstfield JL, Schnall A, Seaquist ER, for the Action to Control Cardiovascular Risk in Diabetes (ACCORD) 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

8. The Diabetes Control and Complications Trial Research Group. The relationship of glycemic exposure (HbA1c) to the risk of development and progression of retinopathy in the Diabetes Control and Complications Trial. Diabetes 1995;44:968–983
9. Stratton IM, Adler AI, Neil HA, Matthews DR, Manley SE, Cull CA, Hadden D, Turner RC, Holman RR. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. BMJ 2000;321:405–412
10. Lachin JM, Genuth S, Nathan DM, Zinman B, Rutledge BN; DCCT/EDIC Research Group. Effect of glycemic exposure on the risk of microvascular complications in the Diabetes Control and Complications Trial—revisited. Diabetes 2008; 57:995–1001
11. Reboussin DM, DeMets DL, Kim KM, Lan KK. Computations for group sequential boundaries using the Lan-DeMets spending function method. Controlled Clinical Trials 2000;21:190–207