CASE 5

Early Termination of the Diabetes Control and Complications Trial

John M. Lachin
Patricia Cleary
Oscar Crofford
Saul Genuth
David Nathan
Charles Clark
Frederick Ferris
Carolyn Siebert
for the DCCT Research Group

ABSTRACT

The Diabetes Control and Complications Trial (1983–1993) of 1,441 subjects followed for an average of 6.5 years assessed the effects of intensive therapy aimed at maintaining near normal levels of blood glucose versus conventional therapy on the risks of diabetes complications of the eyes, kidneys, and nerves. The study was designed to test the hypothesis that the higher than normal blood glucose levels associated with conventional insulin therapy caused these complications. The study was terminated one year ahead of schedule by the monitoring board. This paper describes the medical, ethical, and statistical challenges faced by the study group and the monitoring board.

INTRODUCTION AND BACKGROUND

The Diabetes Control and Complications Trial[1] was a multi-center, randomized, controlled clinical trial of the relative effects of a program of intensive versus conventional management of blood glucose levels on the development and/or progression of microvascular complications of type 1 diabetes mellitus (T1DM). The trial was organized and funded by the National Institute of Diabetes, Digestive, and Kidney Diseases (NIDDK) of the National

Institutes of Health (NIH). The study group was appointed in January 1982, the first subject randomized in August 1983, and the last in June 1989. The study was terminated after an average of 6.5 years of follow-up in June 1993, one year ahead of schedule.

The Director of the NIDDK appointed an external data monitoring committee, called the Data, Safety. and Quality Review Group (DSQRG), to review the accruing data from the trial periodically and to advise on the early termination of the trial or modification of the protocol based on the emerging results. Early in the trial, the DSQRG prepared a document entitled *Operating Procedures for the Data, Safety, and Quality Review Group* which delineated the roles, responsibilities, and functions of the DSQRG. That document was later described in detail by Siebert and Clark.[2] The DSQRG met approximately every six months for the duration of the trial. At the December 1992 meeting, the DSQRG recommended that the DCCT initiate closeout activities as a prelude to consideration of early termination.

The NIDDK had also appointed a second oversight committee, the Policy Advisory Group (PAG), that met periodically to review the continuing viability of the DCCT in light of other emerging evidence while masked to the DCCT results. It was also the responsibility of the PAG to offer a final recommendation on termination of the DCCT when so recommended by the DSQRG. Thus, the analyses of the updated study data were presented to a joint meeting of the DSQRG and PAG in June 1993 at which time both groups concurred that the study should be stopped. The principal results were then rapidly published,[1] followed by dozens of papers on the detailed results of the study. A complete bibliography is available from the website of the Coordinating Center at the George Washington University Biostatistics Center (www.bsc.gwu.edu). The members of the DSQRG and PAG are named in DCCT.[1]

This chapter describes the various considerations which lead to the conclusion by the DSQRG in December 1992 that a statistically significant and clinically meaningful difference between the treatment groups had been observed. Some, but far from all, considerations were statistical. We also describe lessons learned from the monitoring of this trial that may bear on the conduct of future trials.

The Glucose Hypothesis

Type 1 diabetes mellitus (T1DM) is the result of an autoimmune process that leads to ablation of the insulin secreting β-cells of the islet of Langerhans in the pancreas. Eventually the patient decompensates with rising blood glucose levels and other metabolic abnormalities and untreated, eventually dies. In 1922, Banting and Best of the University of Toronto showed that injec-

tions of insulin extracted from animals could lower glucose levels and sustain life.

Within 20 years of the introduction of insulin therapy, a variety of long-term complications of the eyes, kidneys, and nerves (retinopathy, nephropathy, and neuropathy), rarely if ever seen in the pre-insulin era, were observed that ultimately lead to blindness, end stage renal disease, and amputations, respectively. One school of thought postulated that these complications were a manifestation of the underlying course of diabetes per se, or perhaps side effects of exogenous insulin therapy. The other school advocated *The Glucose Hypothesis* that complications resulted from the elevated levels of glycemia (hyperglycemia) that persisted with conventional insulin therapy, and could be prevented by maintaining near-normal levels of glycemia.

The principal weakness of prior studies of this hypothesis was that the technology to achieve and sustain levels of glycemia close to the non-diabetic range simply did not exist. However, by 1980 advances in therapy allowed subjects to achieve near-normal day-to-day levels of glycemia. Multiple daily injections of combinations of short, intermediate and long-acting insulins, or use of a continuous subcutaneous insulin infusion device (or pump), in conjunction with hand-held blood glucose meters, allowed subjects to test their blood glucose levels frequently during the day, before and after meals, and to adjust their insulin doses accordingly. In addition, the glycosylated hemoglobin (HbA_{1c}) assay provided an objective, reliable measure of the average glucose level over the preceding 2–3 months. This provided direct feedback that allowed the clinician and patient to tailor a regimen of diet, exercise, and insulin administration to achieve long-term glucose levels as close to normal as possible. These advances made it practical to conduct a definitive clinical trial to formally test the glucose hypothesis.

PROTOCOL DESIGN

Treatments and Timeline

Since it was impractical to "clamp" a subject at a randomly assigned specific level of glucose, the chosen design assigned half the subjects to receive an intensive therapy aimed at near-normal glycemia, and half to receive conventional therapy with no glucose targets using no more than two insulin injections daily. The principal potential adverse effect of intensive therapy was an increased risk of episodes of hypoglycemia, where low levels of blood glucose cause symptoms ranging from sweating or dizziness to loss of consciousness and seizure. The objective of the study was to evaluate the effects of intensive versus conventional therapy on the risks of retinopathy principally, and also the risks of nephropathy, neuropathy, and hypoglycemia.

The design of the study has been published[3–5] and the protocol, manual of operations, and complete study data sets can be obtained from the National Technical Information Service. Briefly, the study consisted of two independent trials designated as the Primary Prevention Trial and the Secondary Intervention Trial with different eligibility criteria. The primary prevention trial consisted of 726 subjects with early duration type 1 diabetes (1–5 years), no retinopathy, and near-normal renal function (albumin excretion rate (AER) < 40 mg/24 hr). The secondary trial included 715 subjects with longer duration diabetes (1–15 years), minimal background retinopathy, and possibly some early signs of nephropathy (AER < 200 mg/24 hr).

The DCCT study group was organized in January 1982. The study began with the enrollment of 278 subjects into a preliminary trial from August 1983 to March 1984 that demonstrated feasibility.[6] Recruitment to the full-scale study was opened in February 1985 and closed in July 1988 for the secondary trial and June 1989 for the primary trial with an additional 1,163 subjects enrolled, or 1,441 total.[5] All subjects were to be followed through 1993. However, the DSQRG recommended early termination and the final subject visits were held during January–April 1993.

Primary Outcome ~ Retinopathy

The principal DCCT outcome was onset or progression of diabetic retinopathy based on centrally graded fundus photographs obtained from each subject at baseline and at six-month intervals during the trial. Photographs were graded using a 25-step scale of increasing severity of retinopathy in the two eyes (Table 1) that had been developed for the Early Treatment of Diabetic Retinopathy Study.[7] The principal outcome measure was a sustained progression of at least three steps (*sustained 3+ step progression*) from the level on entry (step 1 in the primary trial, steps 2–9 in the secondary) that was observed on two successive six-monthly visits. The principal analysis was specified to be a lifetable analysis of the cumulative

Table 1 Steps of severity of diabetic retinopathy (DR) and levels of severity of diabetic nephropathy

Step	Retinopathy Severity of DR
1	No retinopathy
2–3	Microaneurysms only
4–5	Mild non-proliferative (NPDR)
6–9	Moderate NPDR
10–11	Severe NPDR (SNPDR)
12–25	Proliferative (PDR) and worse

incidence of the onset or progression of retinopathy using a modified Kaplan-Meier estimator and the Mantel-logrank test.[8] Using the method of Lachin and Foulkes,[9] 700 subjects were required to provide 90% power to detect a 37.5% risk reduction, allowing for 10% losses to follow-up and 20% non-compliance, for the primary and secondary trials, or 1,400 total. Power was higher if the rate of loss to follow-up and non-compliance were lower.

However, a 3+ step progression within the above ranges of retinopathy severity is a surrogate outcome that is not usually associated with any lesions or overt symptoms, such as change in vision, requiring treatment. Thus, prior to the start of the full-scale trial, the study investigators recommended to the DSQRG that a treatment group difference in the cumulative incidence of 3+ step progression alone should not be used as a criterion for premature termination of the trial. Rather, they desired that a treatment effect on the incidence of more severe levels of retinopathy be used as the basis for such a decision, such as the incidence of severe non-proliferative diabetic retinopathy or the incidence of laser surgery (photocoagulation). Owing to the lower expected frequency, the protocol specified that treatment group differences for these outcomes would be assessed in the combined primary and secondary trials.

Other Outcomes—Nephropathy and Neuropathy

The DCCT was not designed to detect differences between the treatment groups in the incidence of progression of nephropathy or neuropathy, which occur less frequently than retinopathy. Nevertheless, these and other outcomes were monitored and employed in analyses of the emerging results. Nephropathy outcomes were predefined using the albumin excretion rate (AER). *Microalbuminuria* (or worse) is a value distinctly above normal (AER $\geq$ 40 mg/24 hr), and *albuminuria* (AER $\geq$ 300 mg/24 hr) is equivalent to overt proteinuria, the earliest clinical manifestation of significant diabetic renal disease. Upon entry, subjects were required to have an AER <40 or <200 mg /24 hr in the primary or secondary trial, respectively. It was pre-specified that the cumulative incidence of albuminuria would be assessed in the combined trials due to the expected low incidence.

Autonomic neuropathy was assessed every two years, and neuropathy assessed clinically and by testing of nerve conduction velocity at baseline, five years, and study end. Other outcomes included quality of life, neurocognitive function, mental status, macrovascular events, and risk factors such as blood pressure and serum lipids level. In addition, various adverse effects of diabetes or its treatment were monitored continuously, especially episodes of hypoglycemia. Virtually all outcome assessments were analyzed and monitored periodically by the DSQRG.

The investigators and patients were masked to progression of complications until such time as a level was reached for which treatment was clinically indicated, including severe non-proliferative treatment that could require photocoagulation, renal insufficiency, hypertension, hyperlipidemia, or macrovascular events.

Group Sequential Procedures

The DSQRG and the Coordinating Center jointly specified the statistical procedures for interim monitoring of the accumulating data at approximately six-month intervals. The group-sequential procedure of Lan and DeMets[10] was employed using the "O'Brien-Fleming-like" $\alpha_1^*(t)$ alpha-spending function where t is a measure of the fraction of study information available at a given interim analysis. No formal procedures were applied for monitoring of adverse events (e.g., hypoglycemia) or to monitor for futility (lack of effectiveness).

While these methods provide a stopping boundary, they were employed in a less rigorous way to assess the strength of evidence that a true difference had likely emerged. The DSQRG did not commit itself to terminating the trial if significance was reached for any one analysis, but rather agreed in advance to consider early termination when a body of evidence had emerged that was clinically compelling and that addressed all of the study objectives. The Operating Procedures for the DSQRG described a number of criteria other than statistical significance which should be met prior to any decision to terminate the trial prematurely (see Table 2).

In statistics, "information" has a precise meaning, but for many of the analyses employed in the DCCT, such as lifetables, the precise amount of statistical information to be observed during the entire trial could not be quantified before the end of the trial was reached. Thus, as later described by Lan and DeMets,[11] a function of the duration of the trial was used as a surrogate measure of information. The DSQRG met in November1985 to monitor for the first time both treatment effectiveness and safety. Since close-out was scheduled to occur at the end 1993, 17 semi-annual meetings of the DSQRG were anticipated through December 1993. The fraction of DSQRG meetings held was employed, as a surrogate measure of information. Thus, at each meeting, the Lan–DeMets spending function was employed, with an increment in information of $1/17 = 0.059$.

Longitudinal analyses of repeated measurements over time were also performed using the multivariate rank test of Wei and Lachin[12] that provides a single test of the average difference between treatment groups over all repeated visits combined.[13,14] This method was employed to assess group differences in the distributions of the ordinal retinopathy severity scores over

Table 2 DSQRG Considerations for Early Termination

Excerpted from *Operating Procedures for the Data, Safety and Quality Review Group*.

a. Whether the magnitude or character of an observed difference constitutes a clinically important benefit or risk;
b. Whether the results could be explained by possible differences in baseline variables between the groups;
c. Whether the results could be due to ascertainment bias caused by differences in the treatment regimens;
d. Whether the results are consistent with those for other variables which should be associated with the variable in question;
e. Whether the results are consistent among various subgroups of subjects and across the various centers involved in the study;
f. Whether the risk which is under consideration is outweighed by assessment of the overall potential benefit of therapy;
g. Whether the results could be due to concomitant therapy not directed at blood glucose control rather than due to the different treatment regimens;
h. Whether it is likely that the current trends in the data could be reversed if the trial were to be continued unmodified;
i. Whether and how much additional precision or certainty in the results could be obtained by continuing the trial under the present Protocol; and,
j. Whether there would be significant loss in external validity or credibility of the trial by change in Protocol or discontinuation.

time, longitudinal measures of renal function with severely skewed distributions, and measures of nerve conduction for which there is a lower limit of quantification. These analyses were implemented at the 11th meeting of the DSQRG in December 1990 at an information fraction of 0.647, with subsequent increments of 0.059 in the study information as for the lifetable analyses. For each outcome variable, the correlations of successive test statistics from each DSQRG meeting were computed using the methods described in Su and Lachin.[15] The critical values then were computed by numerical multivariate integration or Monte Carlo simulation.

For more serious but less frequent outcomes such as severe nonproliferative diabetic retinopathy and albuminuria, it was decided to employ a nominal significance level of 0.05 at the end of the trial. For such outcomes, criteria for statistical significance alone were considered less important than the observation of a biologically consistent treatment group effect in conjunction with an effect on 3+ step progression that met group sequential criteria for significance.

In order to maximize the scientific gain from procedures performed infrequently (e.g., nerve conduction studies), the DCCT protocol included a "study-end" evaluation of all DCCT subjects. Thus the study adopted a *Close-out Protocol* which called for a staged termination of the trial if warranted.

The DSQRG would first decide that the trial should initiate closure activities. Thereafter, all subjects would be comprehensively evaluated over a five-month period. These data would then be analyzed and presented to a joint meeting of the DSQRG and PAG at which a decision would be reached regarding the termination of the trial.

THE DATA MONITORING EXPERIENCE

The following is a synopsis of the principal findings at the meetings of the DSQRG leading up to the decision to terminate the study ahead of schedule. A summary of the levels of significance of the principal analyses is presented in Table 3. Only the lifetable analyses of retinopathy and nephropathy progression, and the Wei-Lachin point-prevalence analyses of neuropathy at five years of follow-up are presented.

Figure 1 presents the cumulative incidence of sustained 3+ step retinopathy progression in the primary and secondary cohorts, from DCCT.[1] At the conclusion of the study in June 1993, some subjects had been followed for nine years, with an average of 6.5 years. During the early part of the study there was much discussion of the lack of evidence of benefit. In both cohorts, there was no discernible benefit of intensive therapy between groups during the first five years of follow-up. In the secondary intervention cohort, the risk of "early worsening" during the first two years was increased somewhat with intensive therapy. Nevertheless, intensive therapy continued to yield mean-

Table 3 Emergent Significant Results (P-values and Relative Risk Estimates)

	P-values				Relative risk
	6/91	12/91	6/92	12/92	(95% CI) 12/92
Lifetable analyses					
Boundary (p)	0.011	0.015	0.019	0.024	
Sustained 3+ Step change					
Primary trial	NS	0.004	<0.0001	<0.0001	3.7 (2.2, 6.2)
Secondary trial	0.005	0.0006	<0.0001	<0.0001	2.2 (1.6, 2.9)
Total	0.004	<0.0001	<0.0001	<0.0001	2.5 (1.9, 3.3)
SNPDR	NS	NS	0.048	0.044	1.6 (1.0, 2.6)
Photocoagulation	0.027	0.037	NS	0.013	2.9 (1.2, 6.8)
Albuminuria	NS	NS	0.035	0.016	2.1 (1.1, 3.8)
Prevalence analysis					
Boundary (p)	0.0005	0.0005	0.0013	0.0067	
Neuropathy at 5 y	0.0009	0.00045	<0.0001	<0.0001	2.3 (1.3, 5.4)

Results nominally significant with a p-value (0.05 are shown; NS = not nominally significant). Group sequential boundary critical p-values at the 0.05 significance level (two-sided) are also shown.

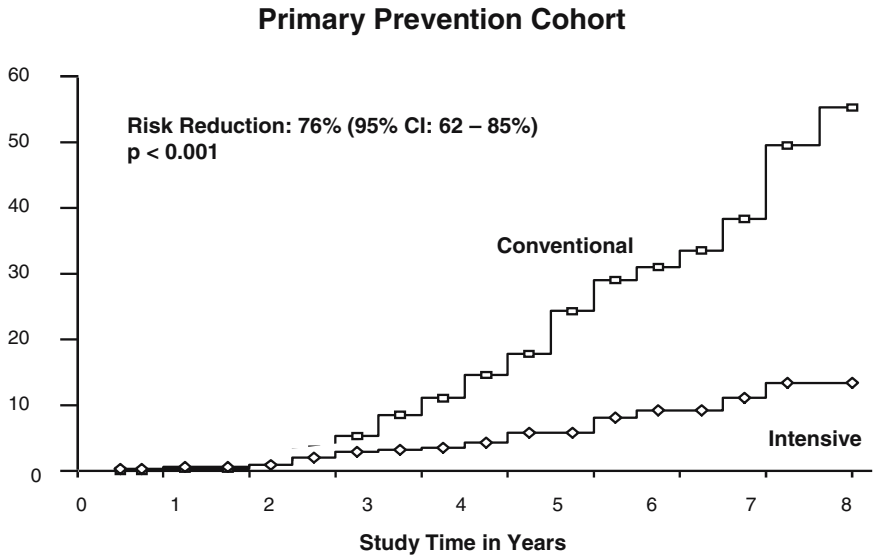

Figure 1 Cumulative incidence of sustained progression of 3 or more steps on the ETDRS scale of retinopathy severity separately within the DCCT primary prevention and secondary intervention cohorts, with the associated risk (hazard) reduction for intensive versus conventional therapy. Reproduced from DCCT (1993) with permission of the *N. Engl J Med*.

ingfully lower levels of blood glucose (HbA_{1c}) and a constant three-fold greater risk of hypoglycemia, as expected; there were no clinically significant increased risks of adverse outcomes with intensive therapy. Accordingly the DSQRG recommended that the trial continue.

In June 1991, at the 12th interim analysis, the lifetable analysis of the incidence of a sustained 3+ step progression within the secondary, but not the primary, trial reached group sequential significance. The analysis of clinically significant neuropathy at five years was nominally statistically significant but did not meet the group sequential criterion for significance. The DSQRG did not find these data to be compelling and recommended that the trial be continued. However, additional analyses were requested, some based on the criteria specified in Table 2.

In December 1991 group sequential significance was observed in the lifetable analysis of a sustained 3+ step progression within both the primary and secondary trials, and in the prevalence of clinically significant neuropathy at five years of follow-up. A nominally statistically significant difference was observed in the lifetable analysis of photocoagulation among all subjects combined.

At this meeting, a variety of additional analyses were presented. The first concerned the patterns of events leading up to the emergence of the significant difference in the lifetable analysis of sustained 3+ step progression. It was determined that the increase in the number of events observed in recent meetings could not be explained by any methodologic factors and was largely due to the increasing accumulation of subject-years of exposure.

Another analysis showed that the observation of a single 3+ step progression at any one visit was associated with an 8.6-fold increase in the risk of developing severe non-proliferative diabetic retinopathy at a future visit (95% confidence limits: 2.7, 14.5) during the study; and a sustained 3+ step progression with a 13-fold increase in this risk (95% CI: 2.5, 23.3). This analysis, therefore, confirmed the predictive importance of 3+ step progression. However, the treatment effect on the risk of severe non-proliferative diabetic retinopathy itself was not statistically significant.

For the first time, the DSQRG entertained a serious discussion of the potential for early termination of the trial. The consensus was that there was a conclusive reduction in risk of the principal outcome, a sustained 3+ step progression in retinopathy, with intensive versus conventional therapy within both the primary and the secondary trials. However, these results alone were not considered clinically compelling because they would not provide a sound basis for treatment recommendations. Therefore, the DSQRG concluded that the study should be continued, but also asked that the Coordinating Center initiate more extensive analyses of retinopathy to address the additional considerations specified in the *Operating Procedures of the DSQRG* (Table 2).

In June 1992 the differences previously observed in 3+ step retinopathy progression and neuropathy persisted, but that in photocoagulation did not. For the first time, the lifetable analyses of severe non-proliferative diabetic retinopathy and of albuminuria were nominally significant in the combined trial. Additional analyses demonstrated that beneficial effects of intensive therapy on retinopathy progression were observed to some degree within specified subgroups of subjects and that there were no major differences among clinics, and that no one or two clinics accounted for the treatment effect.

The general conclusion of the DSQRG was that these analyses satisfied all the criteria necessary for a clinically meaningful treatment group difference in retinopathy. Nevertheless, the DSQRG did not think that all of the major research questions had been answered and questioned whether the current results would be sufficient to inspire a general change in clinical practice. The DSQRG recommended continuation of the study but requested further analyses of hypoglycemia and other adverse effects to better define the benefit to risk ratio of intensive versus conventional treatment.

In December 1992 the analyses of retinopathy progression and neuropathy were group sequentially significant, and those of other more severe outcomes were nominally significant. Table 3 presents the estimated relative risk for conventional versus intensive treatment. The treatment benefit in risk of retinopathy progression was somewhat greater in the primary than the secondary trials. In the total study, the relative risk was 2.5 with 95% confidence limits (1.9, 3.3). This represents a 60% reduction in risk with intensive treatment (95% limits: 47%, 70%). The lifetable analysis of more severe and clinically significant levels of retinopathy and neuropathy also achieved nominal significance within the secondary trial and for both trials combined. In each case, there were too few events within the primary trial to achieve significance, but the observed relative risk was comparable to that within the secondary trial.

Additional analyses of nephropathy demonstrated that beneficial effects of intensive therapy were observed to some degree within subgroups of subjects and that there were no major differences among clinics. Additional analyses also demonstrated that the three-fold increase in the risk of severe hypoglycemia with intensive versus conventional therapy persisted over the full duration of follow-up, was present more or less in all subgroups of the cohort, was relatively stable over time, and was inversely related to the mean HbA_{1c} in both groups.

Overall, therefore, both trials provided strong evidence of clinically meaningful benefit with intensive treatment, and all of the criteria specified in Table 2 were addressed and satisfied. Accordingly, the DSQRG voted unanimously to recommend that the DCCT initiate close-out procedures. At a subsequent meeting in June 1993, based on a preliminary final data set, the DSQRG and Policy Advisory Group jointly recommended that the trial be terminated. Those preliminary results were presented at the national meeting of the American Diabetes Association within weeks of this decision. The final data set was subsequently closed and the major results published in the three months after the final decision to terminate the trial.[2]

LESSONS LEARNED

Methodological Research

One of the major lessons from the DCCT is that the Coordinating Center should be funded to conduct methodological research to address issues posed by the study. The DCCT started with a feasibility trial with a sample size of 278 determined for the analysis of a feasibility outcome. Lachin and Foulkes[9] describe procedures for sample size evaluation for the Mantel-logrank test that allowed for stratification and losses to follow-up, and that

provided the total target sample size of 1,400 necessary for the full-scale DCCT given the initial feasibility study sample size.

The treatment assignments in the DCCT were unmasked, thus admitting the potential for selection and experimental biases. Lachin[16] and Wei and Lachin[17] describe the statistical properties of randomization procedures in general, and Wei's urn randomization procedure, respectively. Based on this and other research, the urn procedure was selected for the DCCT randomization to minimize these biases.

For the longitudinal analysis of the ordinal retinopathy scores in the DCCT, and other measurements, Wei and Lachin[12] developed a family of multivariate rank tests. This approach was further generalized by Thall and Lachin[13] and Lachin.[14] Su and Lachin[15] then described a group-sequential procedure for the Wei-Lachin multivariate rank test that was employed in the interim analyses of the DCCT.

The distribution of rates of hypoglycemia had an excess of zeros and a long tail, relative to a Poisson distribution. Bautista, Lan, and Lachin explored methods for the analysis of such over-dispersed count data. Chapter 8 of Lachin[18] describes the method that was employed for the analyses of hypoglycemia and other event rates in the interim and final analyses of the study data.

The group sequential boundary for the primary outcome had been crossed many times before the study was terminated. However, the group sequential critical values were not used in the publication of the final results.[1] Rather, all results were cited as "nominally significant" at $p \leq 0.05$ (two-sided). Lan, Lachin, and Bautista[19] showed that if the boundary is crossed but the trial continues, then it is conservative simply to employ the fixed sample size critical values in the final analyses, as done in DCCT.[1]

Other Lessons

There were many other lessons from the DCCT of a more practical nature.

When planning the study, diabetic retinopathy was selected as the primary outcome because previous studies had demonstrated that it could be reliably assessed and that it was a highly sensitive measure of retinal abnormalities. While no prior study had used 3+ step progression as an outcome, duplicate gradings had shown that this level of progression was highly reproducible, sensitive, and specific. Further, longitudinal epidemiologic studies had provided a basis for estimation of the expected hazard rate in the conventional group that formed the basis for the sample size evaluation for the study. As it turned out, this estimate was too high. This underscores the importance of sound epidemiologic data for the natural history of the primary outcome in the planned population in designing a clinical

trial, and for being conservative in the assessment of sample size when there is uncertainty regarding the available data.

The power of a study of incidence (time-to-event) is a function of the number of events observed; that in turn is a function of sample size and study duration. From preliminary studies it was estimated that the median time to retinopathy progression was 3.5 years. In order to ensure that any difference in cumulative incidence documented by the trial could reasonably apply to the entire cumulative incidence curve, ranging up to at least the 75th percentile, the study was designed to have an average duration of follow-up of at least seven years. This precaution proved fortuitous since no difference in risk was observed over the first five years of follow-up.

Recognizing the uncertainty of the estimated hazard rate in the conventional group, there was some concern that losses to follow-up and non-compliance would erode the power of the study. Thus the study group insisted on a conservative assessment of sample size that provided at least 90% power using a two-sided test at the 0.05 level after adjusting for 10% losses to follow-up and 20% non-compliance using the model in Lachin and Foulkes.[9] However, these are adjustments for the loss of information, not the bias that can be introduced by losses or non-compliance. To limit the erosion of power and the introduction of bias, the study was implemented using an *intent-to-treat design* in which all patients are followed to the planned study end regardless of adherence to the assigned therapy or side effects of therapy. Extensive subject education was conducted during the recruitment phase[20] to promote compliance with the assigned treatment and complete follow-up, and no subject was permanently withdrawn from study follow-up. The success was remarkable. Of the 1,441 subjects randomized, 32 were declared temporarily inactive at some point during the study, but most of these later returned to follow-up and their assigned treatment. Only eight of those surviving did not attend a final close-out visit in 1993. During the study, subjects adhered to the assigned treatment for 97% of scheduled visits.

This was fortunate because the hazard rate for the primary outcome in the conventional group was substantially less than that projected—0.05 and 0.07 per year in the primary and secondary trials, respectively, versus a projection of 0.2 in each. Thus the loss of power due to the lower hazard rate was offset by the gains in power due to higher than projected rates of follow-up and compliance.

While the hazard rates of such progression within the two treatment groups were not proportional over time, it would be cheating to assess the pattern of the hazards first and to then select the test that appears to be optimal for that pattern. In fact the hazard increased exponentially in the conventional group, while it remained nearly constant in the intensive group. While the power of the Mantel-logrank test was degraded due to the

non-proportional hazards, it was still more sensitive to such patterns than other possible rank tests, such as the Wilcoxon, and in many respects was robust to departures from this assumption. The combination of the conservative assessment of sample size and a relatively robust statistical test helped to ensure that the trial was not underpowered to detect effects of interest.

One of the most important elements in the successful interim monitoring for the DCCT was the selection of DSQRG and PAG members with expertise in all of the areas relevant to the DCCT. These included adult and pediatric diabetes, endocrinology, ophthalmology, nephrology, neurology, cardiology, neuropsychology, ethics, and biostatistics. While many studies have a single statistician member of a DSMB, in the DCCT it was highly advantageous to have three statistician members with different areas of expertise.

Another important step was the development beforehand of a Manual of Procedures[1] for the operation of the DSQRG that covered all aspects of group responsibilities and functions, with input from the study group. This also included a pre-specification of the statistical monitoring plan. No one can foresee the patterns of data that will be observed in a study. What is important, however, is to try to think through the criteria to be used as the basis for a decision to terminate or modify a trial. To the extent possible supplemental analyses, such as subgroup analyses, should be pre-specified.

Despite all these steps, it took many years for the beneficial effects of intensive therapy to evolve. While some might consider that there was cause to consider termination for futility during the early years, this was not the case. The DSQRG realized that these early looks only represented a minor amount of the planned information to be accrued. While it might have been predicted that the benefits of intensive therapy would become manifest sooner, we now understand that hyperglycemia has long-term pervasive physiologic effects that are neither quickly nor completely erased by the implementation of near-normal glycemia, and likewise that the effects of a period of near-normal glycemia are longlasting.[21,22]

It is interesting to note that early in the DCCT, the DSQRG observed a worsening of retinopathy during the first year or so of treatment among subjects assigned to intensive therapy, principally in the secondary intervention trial, where subjects who entered with micro-aneurysms, the earliest sign of retinopathy, developed somewhat more serious sub-clinical lesions. This so-called "Early Worsening" of early retinopathy in patients where tight glucose control is rapidly implemented had been observed in a previous but much smaller trial. The DSQRG reflected on this observation but recommended continuing the trial. With continued follow-up, this excess risk appeared to

dissipate with time; however, there was no evidence of any benefit of intensive therapy for at least the first four years of follow-up, in either the primary prevention or secondary intervention trials. Had a great deal of emphasis been placed on this early worsening, and lack of benefit, the DCCT might have terminated early for harm or futility, missing one of the major advances in the treatment of type 1 diabetes.

While the PAG played an essential role in the DCCT, there is rarely the need for a separate unmasked DSQRG and a PAG that remains masked until a decision is pending. However, while the DCCT was underway there were reports from many smaller studies, some randomized, and it was important to have an independent body charged with continual assessment of the progress (feasibility) of the trial and its relevance in light of other emerging data.

The setting, operational scope, and complexity of the DCCT may have been very atypical. However, every clinical trial is unique in some respects, and these differences may impact the choice of the approach to be adopted for the interim monitoring of the study. While only some of the lessons from the DCCT might apply to another study, we hope that future trials may benefit from our experience.

ACKNOWLEDGEMENTS

John Lachin was Director and Patricia Cleary Co-Director of the Coordinating Center, Oscar Crofford was Chair and Saul Genuth Vice-Chair of the Study, David Nathan was Editor in Chief, Charles Clark was Chair and Frederick Ferris member of the DSQRG, and Carolyn Siebert was NIDDK DCCT Program Director. The members of the research group are presented in DCCT (1993).

The historical account of the DCCT is based on the work of the study group, including the Coordinating Center staff; the Data, Safety, and Quality Review Group; and the Policy Advisory Group. The members of the study group, the DSQRG. and the PAG are presented in the appendix to the DCCT (2) article. Patricia Cleary served as the Co-Director of the Coordinating Center (CoC) for the study duration. Other CoC statisticians (chronologically) included Max Halperin and K.K. Gordon Lan as senior statistical advisors; James Knoke, Desmond Thompson, and Oliver Bautista as research faculty; Peter Gilbert, David Kenny, ShuPing Lan and Jye-yu Backlund as staff statisticians. L.J. Wei and Peter Thall provided statistical consultation. The statistician members of the DSQRG who provided guidance to the committee and the coordinating center on statistical matters were Gary Cutter, David DeMets, and Anastasios Tsiatis.

REFERENCES

1. DCCT Research Group. 1993. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med* 329:977–986.
2. Siebert C, Clark CM. 1993. Operational and policy considerations of data monitoring in clinical trials: The Diabetes Control and Complications Trial Experience. *Control Clin Trials* 14:30–44.
3. DCCT Research Group. 1986. The Diabetes Control and Complications Trial (DCCT): Design and methodological considerations for the feasibility phase. *Diabetes* 35:530–545.
4. DCCT Research Group. 1988. The Diabetes Control and Complications Trial (DCCT): Update. *Diabetes Spectrum* 1:187–190.
5. DCCT Research Group. 1990. The Diabetes Control and Complications Trial (DCCT): Update. *Diabetes Care* 13:427–433.
6. DCCT Research Group. 1987. The Diabetes Control and Complications Trial (DCCT): Results of the feasibility study (Phase II). *Diabetes Care* 10:1–19.
7. Early Treatment Diabetic Retinopathy Study Research Group. 1991. Grading diabetic retinopathy from stereoscopic color fundus photographs: an extension of the modified Airlie House Classification: ETDRS report No.10. *Ophthalmology* 98:786–806.
8. Lachin JM. 2000. Statistical Considerations in the Intent-to-treat Principle. *Control Clin Trials* 21:167–189.
9. Lachin JM, Foulkes MA. 1986. Evaluation of sample size and power for analyses of survival with allowance for non-uniform subject entry, losses to follow-up, non-compliance and stratification. *Biometrics* 42:507–519.
10. Lan KKG, DeMets DL. 1983. Discrete sequential boundaries for clinical trials. *Biometrika* 70:659–663.
11. Lan KKG, DeMets DL. 1989. Group sequential procedures: Calendar versus information time. *Stat Med* 8:1191–1198.
12. Wei LJ, Lachin JM. 1984. Two-sample asymptotically distribution-free tests for incomplete multivariate observations. *J Am Stat Assoc* 79:653–661.
13. Thall PF, Lachin JM. 1988. Analysis of recurrent events: Nonparametric methods for random interval count data. *J Am Stat Assoc* 83:339–347.
14. Lachin JM. 1992. Some large sample size distribution-free estimators and tests for multivariate partially incomplete observations from two populations. *Stat Med* 11:1151–1170.
15. Su JQ and Lachin JM. 1992. Group sequential distribution-free methods for the analysis of multivariate observations. *Biometrics* 48:1033–1042.
16. Lachin JM. 1988. Statistical properties of randomization in clinical trials. *Control Clin Trials* 9:289–311.
17. Wei LJ, Lachin JM. 1988 Properties of the Urn randomization in clinical trials. *Control Clin Trials* 9:345–364.
18. Lachin JM. 2000. *Biostatistical Methods: The Assessment of Relative Risks*. John Wiley and Sons, New York.
19. Lan KKG, Lachin JM and Bautista OM. 2003. Over-ruling a group sequential boundary–a stopping rule versus a guideline. *Stat Med* 22:3347–3355.
20. DCCT Research Group. 1989. Implementation of a multi-component process to obtain informed consent in the Diabetes Control and Complications Trial. *Control Clin Trials* 10:83–96.
21. The DCCT/EDIC Research Group. 2000. Retinopathy and nephropathy in patients with type 1 diabetes four years after a trial of intensive therapy. *N Engl J Med* 342:381–389.
22. The DCCT/EDIC Research Group. 2002. The effect of intensive therapy on the microvascular complications of type 1 diabetes mellitus. *JAMA* 287:2563–2569.