

## Genetic Studies of Late Diabetic Complications

### The Overlooked Importance of Diabetes Duration Before Complication Onset

John J. Rogus,<sup>1,2,3</sup> James H. Warram,<sup>1,2</sup> and Andrzej S. Krolewski<sup>1,2</sup>

**Genes play a role in many processes underlying late diabetic complications, but efforts to identify genetic variants have produced disappointing and contradictory results. Here, we evaluate whether the study designs and analytic methods commonly being used are optimal for finding susceptibility genes for diabetic complications. We do so by generating plausible genetic models and assessing the performance of case-control and family-based trio study designs. What emerges as a key determinant of success is duration of diabetes. This perspective focuses on duration of diabetes before complication onset and its influence on the ability to detect major and minor gene effects. It does not delve into the distinct effect of duration after complication onset, which can enrich case subjects with genotypes conferring survival advantage. We use clinically diagnosed nephropathy in type 1 diabetes to show how ignoring duration can result in considerable power loss in both case-control and family-based trio designs. We further show how, under certain circumstances, disregard for duration information can paradoxically lead to implicating nonrisk alleles as causative. Our results indicate that problems can be minimized by selecting case subjects with short diabetes duration and, to a lesser extent, control subjects with long duration or, perhaps, by adjusting for duration during analysis. *Diabetes* 51: 1655–1662, 2002**

**T**here is ample evidence that susceptibility to late diabetic complications depends in part on genetic factors (1–5). Nevertheless, genetic association studies have thus far failed to implicate causative variants consistently (6). In some cases, inadequate sample sizes may be contributing to this elusiveness, but a contributing factor may be a failure to recognize the impact that duration of diabetes before the onset of the complication can have on the power of a study. This

perspective focuses on this critical consideration for detecting genes that may influence lifetime risk of developing the complication and, particularly, any gene that modifies when in the course of diabetes a complication is likely to occur.

For this perspective, diabetes duration is defined as the period of time from diabetes onset until complication onset. We illustrate the potential importance of this variable by exploring several genetic models that fit the epidemiologic characteristics of diabetic nephropathy as diagnosed by persistent proteinuria. We concentrate on this stage of diabetic nephropathy because duration-related mortality should play a relatively minor role before the onset of proteinuria (1), thereby allowing us to focus solely on the impact of precomplication duration of diabetes.

Two classes of genetic studies are considered, one based on case-control analysis and the other based on family-based trio analysis. These two designs, the workhorses for evaluating whether specific genetic variants are associated with a disease end point, are emerging as the most commonly used tools for studying the genetics of diabetic complications. Case-control studies are attractive because they do not require identifying families with multiple occurrences of diabetes and because power per individual sampled typically is high. Unfortunately, case-control studies are also susceptible to bias if case and control subjects are not drawn from genetically similar populations. Family-based trio analysis, as proposed by Spielman et al. (7,8), overcomes this problem by eliminating the need for a control group. Most commonly, each trio comprises an affected offspring together with both parents. Studies based on such trios (“affected offspring trios”) are logistically demanding, because parents must be identified and enrolled, and typically less powerful than case-control studies. However, they are more robust in the sense that matching of case and control subjects is not problematic. This follows from the fact that genetic variants (i.e., alleles) in parents serve as a reference set for comparison with the alleles in the offspring. Families with unaffected offspring (“unaffected offspring trios”) can sometimes serve as a useful alternative to affected offspring trios, and we consider this design as well.

In both case-control and family-based trio studies, phenotypic dichotomy is usually assumed, implying, for example, that a case of diabetes with 10 years of duration

From the <sup>1</sup>Research Division, Joslin Diabetes Center, Boston, Massachusetts; <sup>2</sup>Department of Medicine, Harvard Medical School, Boston, Massachusetts; and <sup>3</sup>Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts.

Address correspondence and reprint requests to John Rogus, Section on Genetics & Epidemiology, Joslin Diabetes Center, One Joslin Place, Boston, MA 02215-5397. E-mail: john.rogus@joslin.harvard.edu.

Received for publication 25 September 2001 and accepted in revised form 31 January 2002.

before onset of complication is treated the same as a case of diabetes with 25 years of duration. In a general context, Morton and Collins (9) recommended various ways to define "hypernormal controls," such as excluding unaffected individuals who are still young, but much more can be done in the area of late diabetic complications to improve on the standard yes/no definition of disease. To show how diabetes duration before onset of complication can be useful for this purpose, we adopt the approach that Li and Hsu (10) proposed to study age at onset in affected offspring trio studies and extend their work to accommodate both unaffected offspring trio and case-control designs. By evaluating a number of genetic models consistent with epidemiologic data on the occurrence of proteinuria in type 1 diabetes, we show how power and, in some instances, validity could depend on duration of diabetes before onset of complication. Finally, we suggest several ways to incorporate duration data into genetic association analyses of late diabetic complications.

#### MODELING DURATION IN DIABETIC COMPLICATIONS

Epidemiologic studies have provided much information about the incidence of complications according to diabetes duration. For example, among individuals with type 1 diabetes, the incidence of nephropathy peaks during the second decade of diabetes and declines thereafter, whereas the incidence of retinopathy shows no such decline (1,11). Data are also available to allow stratification of incidence curves by environmental factors such as level of glycemic control [1; unpublished data from Krolewski et al. (1)]. To show how diabetes duration may influence the ability of each study design to detect genetic association, we will need to model a similar stratification by genotype (the two-allele combination carried by a given person). The variety of possible genetic models is endless, ranging from a single major genetic effect whereby carriage/noncarriage of a risk allele essentially dictates who will become affected (see Fig. 1A) to a subtle minor genetic effect that simply shortens or lengthens the duration at which onset occurs (see Fig. 1D). (We hasten to note that, although convenient, the terms major and minor are highly subjective.) Given the cumulative incidence rates by duration for either of these extreme models or for any model in between, one may calculate the power to detect genetic association using the methodology developed by Li and Hsu (10). Specifically, formulas for calculating power to detect excess transmission of the risk allele from parents to affected offspring (7,8) have already been worked out (10), and they can easily be modified (by substituting survival functions for density functions) to accommodate trios with unaffected offspring, an alternative suitable for the study of late diabetic complications (12). In the Appendix, we develop appropriate equations for calculating power in case-control studies.

The critical remaining task, then, is to develop appropriate models of cumulative incidence by duration and genotype. Li and Hsu (10) suggested a convenient way to accomplish this by modeling the hazard rate, the instantaneous risk of disease at a point in time given disease-free survival until that point. For simplicity, we assume that the baseline hazard, applicable if no risk alleles are present, follows a Weibull distribution and that hazard functions

for all other genotypes are proportional to the baseline level. Within this framework, there is considerable flexibility to model genetic effects, and, by choosing appropriate shape and scale parameter values, we were able to mimic cumulative incidence among those with type 1 diabetes.

#### GENETIC MODELS CONSIDERED

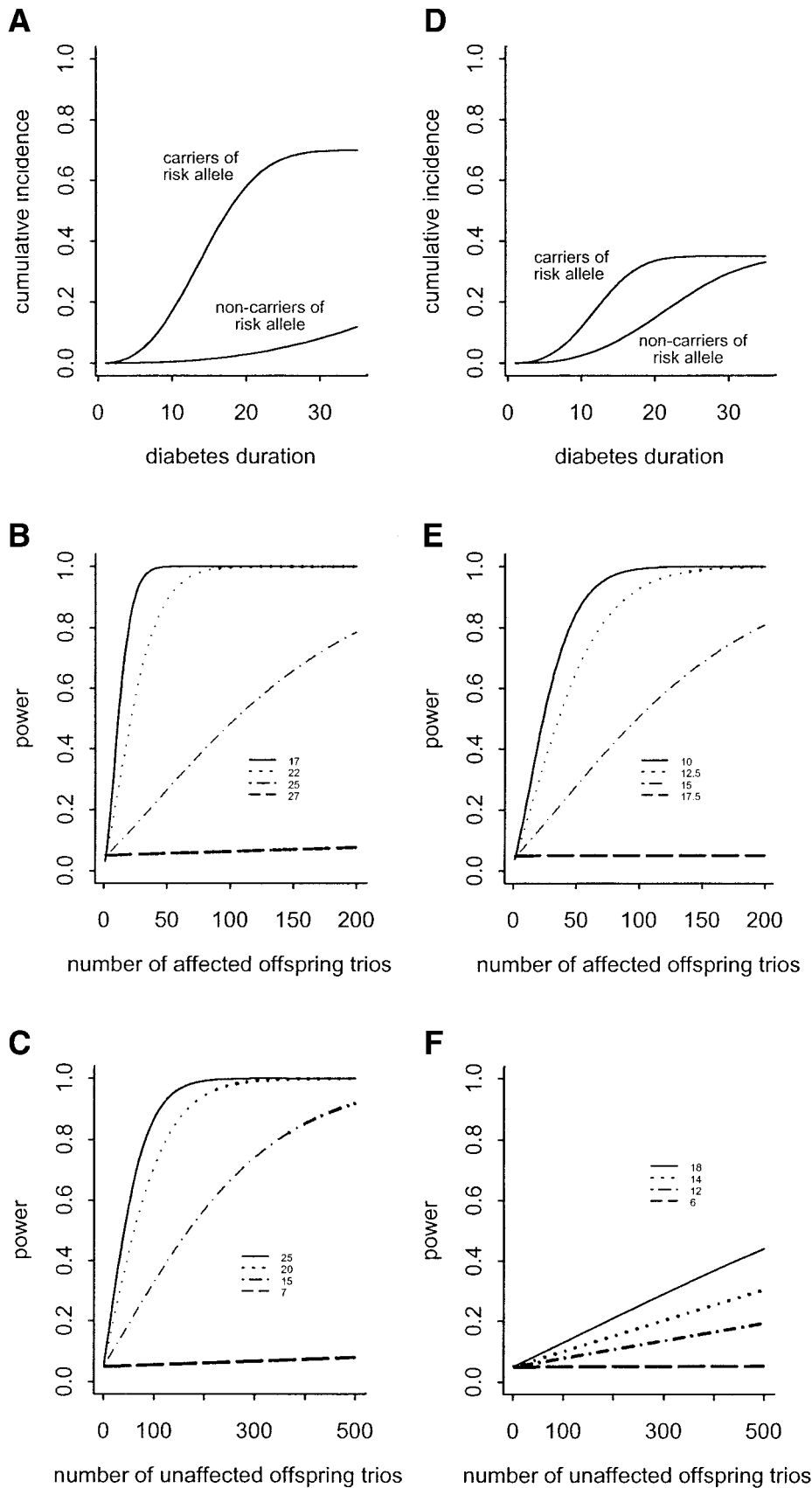
Our illustrative example is based on the occurrence of persistent proteinuria in type 1 diabetes. Epidemiologic studies show that the incidence rate of proteinuria increases rapidly between the 5th and 15th years of diabetes and then declines thereafter, leading ultimately to a lifetime risk of 35% (1). This pattern suggests that a substantial subset of patients are at high risk to develop proteinuria very early in the course of diabetes. Various scenarios are consistent with this finding, but one also consistent with the high sibling recurrence risk reported by Quinn et al. (3) in individuals with type 1 diabetes and with results of complex segregation analysis in individuals with type 2 diabetes (13,14) is a gene imparting a major effect. Genes that play a less pronounced role (minor genes, for lack of better terminology) may also play a role in susceptibility (6).

To explore these two extremes, we consider two dominant acting genes, one with a major impact on susceptibility and one with a minor effect. Parameters are chosen so that carriers of the major gene risk allele have a lifetime risk of persistent proteinuria of 70% compared with 12% for noncarriers (Fig. 1A). Carriers also tend to develop disease more quickly than noncarriers. In contrast, the minor gene model considered here acts primarily by accelerating disease onset. Consequently, both carriers and noncarriers of the minor gene risk allele have a lifetime risk of ~35% (Fig. 1D). In both cases, we assume risk allele frequency of 20%, although other frequencies are considered in sensitivity analyses along with other modes of inheritance. Environmental factors such as glycemic control and other genetic effects need not be explicitly modeled in this approach.

#### EFFECT OF DURATION ON FAMILY-BASED TRIO STUDIES

Any set of family-based trios collected for genetic analysis will contain offspring with varying degrees of diabetes duration. Although it theoretically is possible to calculate power based on the distribution of duration among such a heterogeneous group, it will be more illustrative for us to consider artificial data sets that are composed entirely of offspring with the same level of duration. We use the same approach in the section that deals with case-control studies.

For the major gene model (Fig. 1A), power to detect excess transmission of the risk allele to affected offspring is inversely correlated with diabetes duration before onset of proteinuria. Power curves for samples of case subjects with 17, 22, 25, and 27 years of diabetes duration before onset of proteinuria illustrate this point (Fig. 1B). Although not much efficiency is lost as duration increases from 17 to 22 years, substantial power loss does occur by 25 years of duration. By 27 years of duration, transmission of the risk allele is essentially 50%, resulting in basically no power at all. For the minor gene model (Fig. 1D), a similar



**FIG. 1.** Cumulative incidence of persistent proteinuria according to duration of diabetes in carriers and noncarriers of risk allele for major (A) and minor (D) gene model. Power to detect excess transmission of risk allele to offspring with proteinuria from heterozygous parents for major (B) and minor (E) gene model. Power to detect excess transmission of nonrisk allele to offspring without proteinuria from heterozygous parents for major (C) and minor (F) gene model.

pattern emerges (Fig. 1E), except that substantial power loss begins at ~15 years of duration and accelerates until 17.5 years of duration, when power vanishes completely.

As duration reaches even higher levels, the risk allele is actually transmitted to affected offspring *less* frequently than the nonrisk allele. Amazingly, this deviation from

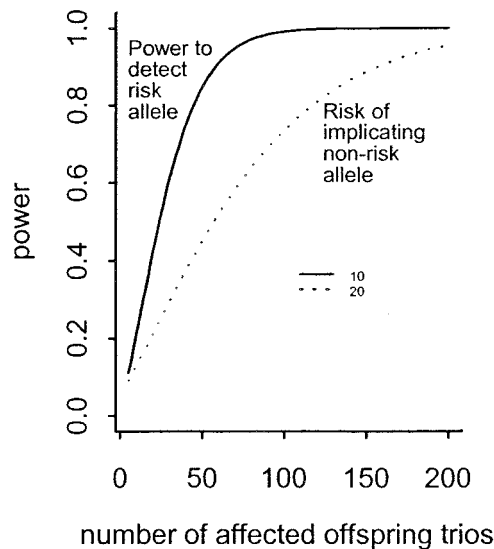


FIG. 2. Power to detect excess transmission of risk allele from heterozygous parents to offspring who developed proteinuria after 10 years of diabetes, and excess transmission of nonrisk allele from heterozygous parents to offspring who developed proteinuria after 20 years of diabetes (minor gene model).

expected transmission can reach statistical significance with samples of moderate size if a two-sided test is used. Figure 2 demonstrates this for the minor gene model by showing how power to detect excess transmission of the *nonrisk* allele using affected offspring trios with 20 years duration is only slightly lower than the power to detect excess transmission of the *risk* allele using affected offspring trios with 10 years duration. Therefore, a sample of case subjects with onset of proteinuria occurring after long duration of diabetes would misidentify the nonrisk allele as being causative. A combined sample of case subjects with short and long duration of diabetes may have basically no power because of the mixture of transmission rates.

Testing transmission rates of risk alleles in trios that comprise an affected offspring and both parents is the most commonly used family-based design, but other options exist. One alternative that may be particularly relevant for late diabetic complications (12) involves evaluating rates at which nonrisk or even protective alleles are transmitted from parents to unaffected offspring (those with diabetes but without proteinuria in our example). Not surprisingly, the degree of allelic transmission and, hence, power depends on duration of diabetes in the unaffected offspring. For the major gene model of proteinuria, power is positively correlated with duration (Fig. 1C). For the minor gene model, power increases slowly with duration until 18 years (Fig. 1F), at which point, to complicate matters, power begins to decrease with duration (data not shown). This reversal, however, is basically academic, because power is uniformly low for detecting nonrisk allele transmission in the minor gene model of proteinuria. Thus, because only genes of large or moderately large effect will be detectable for proteinuria, our results in toto indicate that control subjects with long duration are preferable.

We readily acknowledge that the findings for both types of trios depend in part on the characteristics of the genetic models evaluated, and we are by no means advocating

universal duration-specific guidelines for defining ideal trios. Nevertheless, consistency of the basic trends for many permutations of disease allele frequency and mode of inheritance (results not shown) indicate that, as a general rule, affected offspring trio studies gain efficiency by selecting affected offspring with short duration before onset of complication. In contrast, unaffected offspring trio studies can often benefit by requiring the unaffected offspring to have long-duration diabetes. On a broader level, however, these results highlight the need to be cognizant of diabetes duration before onset of proteinuria or other late diabetic complications in family-based trio studies.

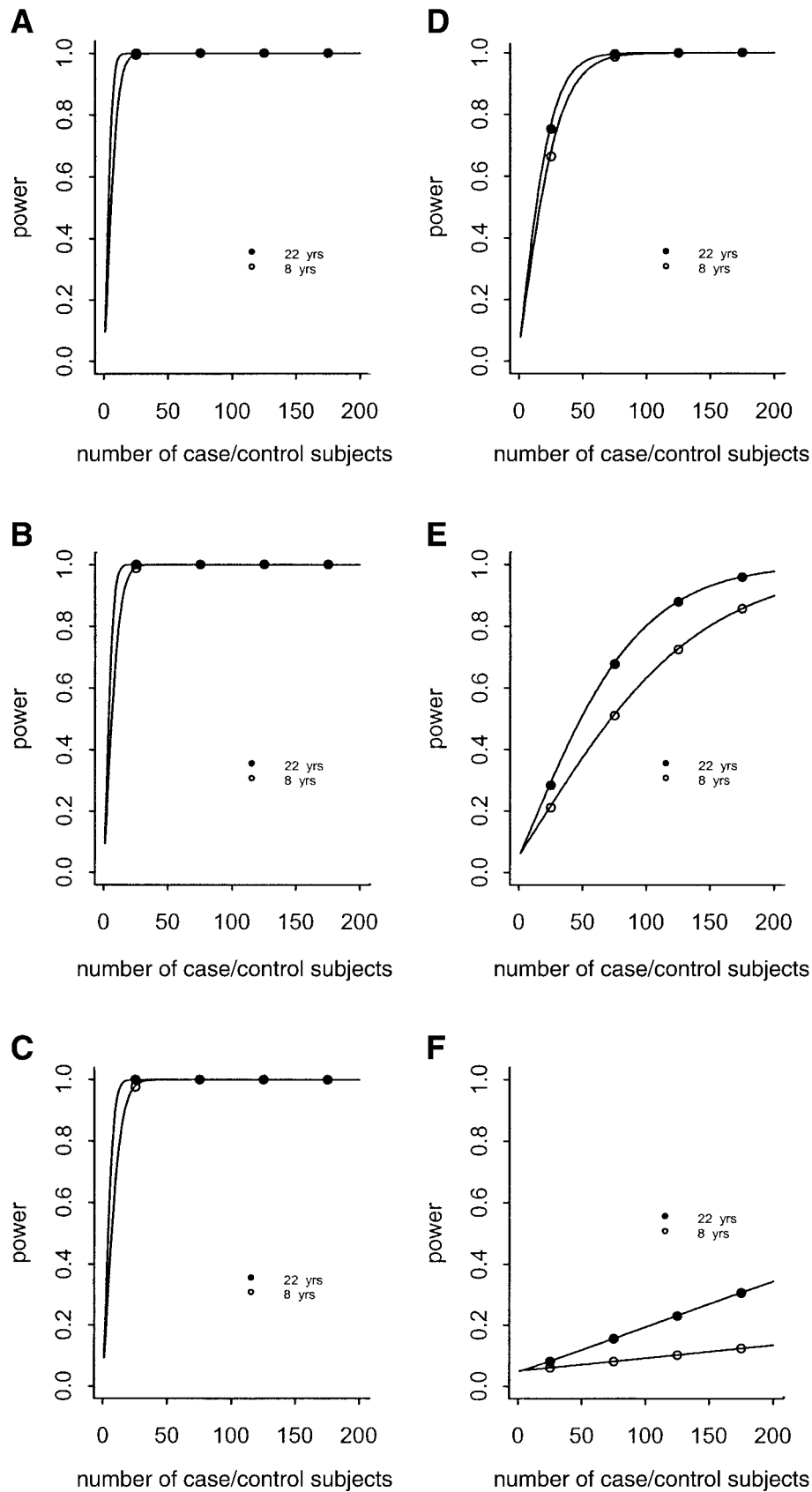
#### EFFECT OF DURATION ON CASE-CONTROL STUDIES

Unlike the family-based trio studies just considered, case-control studies typically compare allele or genotype frequencies of case subjects to those in control subjects. In theory, therefore, power derives from both an increased frequency of risk alleles among case subjects and a decreased frequency among control subjects. Furthermore, within each group, there is the effect of duration that is the explicit focus of our considerations.

Extrapolation of the findings from the previous section suggests that case-control studies should focus on short-duration case subjects and long-duration control subjects. In general, this turns out to be true for our illustrative models, but there is a slight twist related to the fact that case and control subjects are being considered simultaneously. Specifically, given an equal number of case and control subjects, power is not highly sensitive to duration among control subjects. For example, if all case subjects have 10 years of duration, then power under the major gene model is virtually unchanged regardless of whether control subjects have 8 or 22 years of duration (Fig. 3A). This follows because excessive carriage of the risk allele among the case subjects is driving power far more than deficient carriage among the control subjects. Because this excess carriage dissipates slowly with duration among case subjects for the major gene scenario, similar results are found even as case duration increases to 15 (Fig. 3B) or 17 (Fig. 3C) years. Similarly, in the minor gene scenario, control duration plays a very small role when 10-year-duration case subjects (Fig. 3D) are considered and only a slightly more pronounced role for 15-year-duration case subjects (Fig. 3E). What matters more in the minor gene case is the duration among case subjects, as illustrated by considering Fig. 3D–F. Analogous to our findings from the family-based trio models, case subjects with duration longer than some model-dependent value will actually tend to carry nonrisk alleles in excess. For sufficiently long duration, this tendency can result in a higher, possibly significant, frequency of nonrisk alleles in case compared with control subjects.

#### TREATMENT OF DURATION IN RECENT DIABETES ARTICLES

In the past few years, several association studies of late diabetic complications have been published in *Diabetes*. In surveying a recent 2-year period (July 1999 to July 2001), we found wide variability with respect to the quality of duration data, the analytic approaches to handle duration



**FIG. 3.** Power of case-control analysis to detect major (A) or minor (D) gene when case subjects developed proteinuria after 10 years of diabetes and control subjects without proteinuria had diabetes duration of 8 or 22 years. Power of case-control analysis to detect major (B) or minor (E) gene when case subjects developed proteinuria after 15 years of duration of diabetes and when control subjects without proteinuria had diabetes duration of 8 or 22 years. Power of case-control analysis to detect major (C) or minor (F) gene when case subjects developed proteinuria after 17 years of duration of diabetes and when control subjects without proteinuria had diabetes duration of 8 or 22 years.

data, and the discussion of how duration characteristics of the study population may have influenced results (15–20).

In terms of data quality, some studies reported duration at examination, some reported duration at onset of diabetic complications, and some did not comment on dura-

tion at all. Moreover, descriptions of duration estimation procedures were sparse. Many of the details concerning the assessment of diabetes onset (i.e., duration start date) and complication onset (i.e., duration end date) were omitted, making it difficult to gauge the precision of duration estimates.

Two basic analytic strategies were used to deal with duration. The first involved restricting entry of control subjects (but not case subjects) on the basis of duration. In most situations, this meant requiring control subjects to have long duration, but an alternative, choosing control subjects with case-matched duration, was also used. The second strategy involved incorporating duration information at the stage of analysis. In some studies, duration was treated as an independent variable. Variations included using duration as a stratification variable for case-control analysis (e.g., late-onset case versus long-duration control subjects and early-onset case versus short-duration control subjects) and using duration as an explanatory variable in logistic regression. Elsewhere, duration was treated as a dependent variable for testing whether certain genotypes corresponded to longer/shorter average duration. No studies attempted to use survival analysis to determine whether duration until complication varied among genotype groups.

Discussion of the possible impact of duration on results occurred primarily when inclusion of duration strengthened the significance of results. It was not uncommon for duration among case subjects to be longer than duration among control subjects, but this fact was not highlighted even when negative results could have been due in part to this aspect of the data.

#### IMPLICATIONS FOR FUTURE STUDIES

Our abridged literature review demonstrates that there is currently no clear consensus on dealing with duration of diabetes in genetic studies of complications. Moreover, the constant search for better designs and methods virtually guarantees a degree of variability in approaches moving forward. Nevertheless, we believe that it is worthwhile to provide some general comments that may pertain to future studies.

First, more attention could be paid to the quality of duration information. It would be useful if authors reported summary measures of duration among case and control subjects as well as the protocols for obtaining this information. Relevant topics would include methods for ascertaining diabetes onset and complication onset as well as any duration-related exclusion/inclusion criteria. Such information would provide important context for accompanying results and conclusions and would facilitate more meaningful comparisons across studies.

Second, when reliable data are available, researchers could explore whether it is possible to use duration to improve power and/or reduce the potential for bias. At ascertainment, a reasonable, albeit imperfect, rule of thumb is to focus on early-onset case and long-duration control subjects. Admittedly, implementation of this simple idea is complicated by the fact that optimal duration cut-offs are dependent on unknown underlying models. Nevertheless, excluding at least some proportion of late-onset case and short-duration control subjects could have

a dramatic impact. For instance, although the ideal affected offspring trios for our minor gene model would have no more than ~12 years of duration before onset of proteinuria, misjudging this cut-off point by even 5 years would still result in exclusion of the most counterproductive trios, those in which the nonrisk allele is expected to be transmitted preferentially.

Flexible definitions could help alleviate some of the burden imposed by a restrictive ascertainment scheme. Normoalbuminuria after 15 years of diabetes, for instance, may be a sensible alternative to an entry criterion requiring 20 years without proteinuria (and it may also lessen the impact of mortality due to other complications). Moreover, a somewhat less restrictive duration cut-off for control subjects may be acceptable in case-control studies, because excess carriage of risk alleles among case subjects will likely be the primary determinant of power. When collecting short-duration case subjects, extra care should be taken to rule out kidney disease not related to diabetes.

Restricting ascertainment to early-onset case and long-duration control subjects is a simple but perhaps not optimal way to incorporate information on duration. This approach assumes that later-onset case subjects and control subjects with short-duration diabetes are dispensable, and this may not be so in all situations. In fact, there are plausible genetic models that fit the reported incidence data on proteinuria in which duration is irrelevant for either case or control subjects (although probably not both simultaneously). Moreover, the ability to address other important research areas, such as determining how genes interact with drug-based interventions to delay disease onset, may depend on availability of case subjects with late as well as early onset. Therefore, a more sensible approach may be to use analytic methods that are able to accommodate duration data. One easy alternative, stratification by duration group, can be carried out to either produce descriptive data or test for trends with increasing/decreasing duration. More sophisticated approaches include conditional logistic regression with duration as an independent variable or survival analysis. Mokliatchouk et al. (21) provided a detailed discussion of these two methods as they apply to both family-based trio studies and case-control studies. Major considerations in choosing an appropriate method include the type of duration data available (duration-until-onset versus duration-at-ascertainment) and the ascertainment scheme used (population-based versus trait-based sampling). An additional benefit of these statistical approaches is that other covariates such as sex, parental blood pressure, and level of glycemic control can be easily incorporated. Known genes can also be accommodated in a similar manner, as Mokliatchouk et al. (21) demonstrated using an example based on Alzheimer's disease.

Our final comment pertains to interpreting results in a way that thoughtfully considers the potential impact of duration. For some studies, this may involve entertaining the idea that negative or positive results could be due in part to suboptimal duration profiles of study participants (e.g., duration being too long in case subjects).

## OTHER CONSIDERATIONS AND FUTURE RESEARCH

Although this perspective has focused on genetic association studies, it is worth mentioning that linkage studies are also vulnerable to duration-before-onset effects. According to Li and Hsu (10), affected sibling pair studies should focus primarily on siblings with similar age at onset to avoid power loss and potentially biased results. Adapting this finding to late diabetic complications, affected sibling pairs with similar diabetes duration should be best, but additional work could clarify this issue and also provide guidance for discordant sibling pair studies (22).

Future work could also address special considerations relevant to particular aspects of late diabetic complications. For example, the extremely high lifetime risk of proliferative retinopathy among those with type 1 diabetes (11) may increase the importance of long diabetes duration in control subjects for case-control studies. It would also be instructive to examine all of the above issues on each of the successive stages of any given complication. For diabetic nephropathy, this would range from the onset of microalbuminuria to progression from proteinuria to end-stage renal disease. Toward the later stages of disease, this would necessarily involve a detailed look into how increased mortality as a result of end-stage renal disease or cardiovascular disease could affect genotype distributions for genes involved in disease susceptibility as well as genes involved in modifying survival. Failure to appreciate such effects could result in false-positive or false-negative results. Finally, all of these topics must be reconsidered specifically in the context of type 2 diabetes, for which issues such as pinpointing diabetes onset will assume greater prominence.

In addition to incorporation of duration information, Morton and Collins (9) suggested several other ways to improve the efficiency of case and control subjects, and some of these may be relevant for late diabetic complications. For instance, there may be a benefit to choosing case subjects with a positive family history and control subjects with a heavy environmental load (e.g., very poor glycemic control). The Morton and Collins article also goes to great lengths to compare the relative efficiency of case-control and family-based trio studies (9). In general, our results echo their sentiment that case-control studies are more powerful (compare Fig. 1 with Fig. 3), but it would be interesting to pursue this point further in light of other factors, such as survival bias, which, if pertinent in parents, could partially determine which families are available for family-based trio analysis.

The coming years promise to be exciting ones in the field of genetics of diabetic complications. Many laboratories throughout the world are taking part in the search for susceptibility genes, and several large initiatives are currently under way to establish large data resources for genetic studies. These activities will provide a tremendous opportunity to improve our understanding of the genetic basis of kidney disease, heart disease, and eye disease among those with diabetes. Without proper attention to issues such as duration of diabetes, the full potential of these opportunities may not be realized.

## APPENDIX

To calculate power for case-control analysis, we begin by defining  $s(x)$  as the number of risk alleles carried by a case with duration before onset  $x$  and  $t(x)$  as the number of risk alleles carried by a control with duration at ascertainment  $x$ . Under the null, both  $S = \sum_i s(x_i)$  and  $T = \sum_i t(x_i)$  have expectation  $2np$  and variance  $2np(1-p)$ , where  $p$  is the risk allele frequency, so  $U = S - T$  has mean 0 and variance  $4np(1-p)$ . Under the alternative hypothesis ( $D$  is risk predisposing),  $E(S) = \sum_i E(s(x_i))$  and  $\text{Var}(S) = \sum_i \text{Var}(s(x_i))$ , where:

$$\begin{aligned} E(s(x_i)) &= 2 \text{pr}(DD|x_i) + \text{pr}(Dd|x_i) \\ \text{Var}(s(x_i)) &= 4 \text{pr}(DD|x_i) + \text{pr}(Dd|x_i) - E(s(x_i))^2. \end{aligned}$$

By applying Bayes rule, the conditional probabilities  $\text{pr}(DD|x_i)$  and  $\text{pr}(Dd|x_i)$  can be expressed as functions of the known quantities  $\text{pr}(x_i|DD)$ ,  $\text{pr}(x_i|Dd)$ , and  $\text{pr}(x_i|dd)$ . After applying the same argument to find  $E(T)$  and  $\text{Var}(T)$ , we can then calculate  $E(U) = E(S) - E(T)$  and  $\text{Var}(U) = \text{Var}(S) + \text{Var}(T)$ , and estimate the two-sided power for a significance level of  $\alpha$  test as:

$$1 - \phi\left\{\frac{[(4np(1-p))^{1/2} z_{1-\alpha/2} - E(U)]/\text{Var}(U)^{1/2}}{[(4np(1-p))^{1/2} z_{1-\alpha/2} - E(U)]/\text{Var}(U)^{1/2}}\right\} + \phi\left\{-\frac{[(4np(1-p))^{1/2} z_{1-\alpha/2} - E(U)]/\text{Var}(U)^{1/2}}{[(4np(1-p))^{1/2} z_{1-\alpha/2} - E(U)]/\text{Var}(U)^{1/2}}\right\}.$$

## ACKNOWLEDGMENTS

This research was supported by grant 9-2000-1008 from the Juvenile Diabetes Research Foundation and National Institutes of Health Grants R01-DK41526 and R01-DK53534.

## REFERENCES

- Krolewski AS, Warram JH, Christlieb AR, Busick EJ, Kahn CR: The changing natural history of nephropathy in type 1 diabetes. *Am J Med* 78:785-794, 1985
- Seaquist ER, Goetz FC, Rich SS, Barbosa J: Familial clustering of diabetic kidney disease: evidence for genetic susceptibility to diabetic nephropathy. *N Engl J Med* 320:1161-1165, 1989
- Quinn M, Angelico MC, Warram JH, Krolewski AS: Familial factors determine the development of diabetic nephropathy in patients with IDDM. *Diabetologia* 39:940-945, 1996
- The Diabetes Control and Complications Trial Research Group: Clustering of long-term complications in families with diabetes in the Diabetes Control and Complications Trial. *Diabetes* 46:1829-1839, 1997
- Wagenknecht LE, Bowden DW, Carr JJ, Langefeld CD, Freedman BI, Rich SS: Familial aggregation of coronary artery calcium in families with type 2 diabetes. *Diabetes* 50:861-866, 2001
- Krolewski AS: Genetics of diabetic nephropathy: evidence for major and minor gene effects. *Kidney Int* 55:1582-1596, 1999
- Spielman RS, McGinnis RE, Ewens WJ: Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52:506-516, 1993
- Ewens WJ, Spielman RS: The transmission/disequilibrium test: history, subdivision, and admixture. *Am J Hum Genet* 57:455-464, 1995
- Morton NE, Collins A: Tests and estimates of allelic association in complex inheritance. *Proc Natl Acad Sci U S A* 95:11389-11393, 1998
- Li H, Hsu L: Effects of age at onset on the power of the affected sib pair and transmission/disequilibrium tests. *Ann Hum Genet* 64:239-254, 2000
- Krolewski AS, Warram JH, Rand LI, Christlieb AR, Busick EJ, Kahn CR: Risk of proliferative diabetic retinopathy in juvenile-onset type 1 diabetes: a 40-yr follow-up study. *Diabetes Care* 9:443-452, 1986
- Scott L, Rogus JJ: Using unaffected child trios to test for transmission distortion. *Genet Epidemiol* 19:381-394, 2000
- Fogarty DG, Hanna LS, Wantman M, Warram JH, Krolewski AS, Rich SS: Segregation analysis of urinary albumin excretion in families with type 2 diabetes. *Diabetes* 49:1057-1063, 2000

14. Imperatore G, Knowler WC, Pettitt DJ, Kobes S, Bennett PH, Hanson RL: Segregation analysis of diabetic nephropathy in Pima Indians. *Diabetes* 49:1049–1056, 2000
15. Hudson BI, Stickland MH, Futers TS, Grant PJ: Effects of novel polymorphisms in the RAGE gene on transcriptional regulation and their association with diabetic retinopathy. *Diabetes* 50:1505–1511, 2001
16. Poirier O, Nicaud V, Vionnet N, Raoux S, Tarnow L, Vlassara H, Parving H-H, Cambien F: Polymorphism screening of four genes encoding advanced glycation end-product putative receptors: association study with nephropathy in type 1 diabetic patients. *Diabetes* 50:1214–1218, 2001
17. Vaessen N, Heutink P, Janssen J, Witteman J, Testers L, Hofman A, Lamberts SWJ, Oostra BA, Pols H, van Duijn CM: A polymorphism in the gene for IGF-I: functional properties and risk for type 2 diabetes and myocardial infarction. *Diabetes* 50:637–642, 2001
18. Araki S, Moczulski DK, Hanna L, Scott LJ, Warram JH, Krolewski AS: APOE polymorphisms and the development of diabetic nephropathy in type 1 diabetes: results of case-control and family-based studies. *Diabetes* 49:2190–2195, 2000
19. James RW, Leviev I, Ruiz J, Passa P, Froguel P, Garin M-C: Promoter polymorphism T(–107)C of the paraoxonase PON1 gene is a risk factor for coronary heart disease in type 2 diabetic patients. *Diabetes* 49:1390–1393, 2000
20. Neugebauer S, Baba T, Watanabe T: Association of the nitric oxide synthase gene polymorphism with an increased risk for progression to diabetic nephropathy in type 2 diabetes. *Diabetes* 49:500–503, 2000
21. Mokiatouch O, Blacker D, Rabinowitz D: Association tests for traits with variable age at onset. *Hum Hered* 51:46–53, 2001
22. Rogus JJ, Krolewski AS: Using discordant sib pairs to map loci for qualitative traits with high sibling recurrence risk. *Am J Hum Genet* 59:1376–1381, 1996