Several large clinical trials for the prevention of IDDM in islet cell antibody positive first-degree relatives are planned or underway. The design of these trials rests in part on assumptions about the natural history of autoimmunity during the prediabetic period and on the likely effectiveness of the intervention being tested. At this time, most of the factors that influence the required sample size can only be roughly estimated. Diabetes 42:941-47, 1993

Several clinical trials for the prevention of IDDM are underway in the U.S. (1,2), and a large European-Canadian Nicotinamide Diabetes Intervention Trial is about to begin. That trial will involve the screening of ~22,000 first-degree relatives for ICAs and the randomization of ~500 ICA+ relatives into nicotinamide and placebo arms, with 5 yr of follow-up (E.A.M. Gale, unpublished observations). In contrast, a trial of injected insulin, in which we are participating, will screen ~8000 first-degree relatives. Forty-eight high-risk ICA+ individuals, who have a low FPIR on IVGTT, will be randomized and followed for 4 yr. The insulin trial has a 2 x 2 factorial design with two interventions (intravenous and subcutaneous insulin) at two levels (administered and not administered) resulting in four arms: intravenous placebo, intravenous insulin, intravenous placebo plus subcutaneous insulin, and intravenous insulin plus subcutaneous insulin. A comparison of these two trials reveals that they differ by a factor of >5 in the number of relatives screened for ICA (per study arm) and by a factor of 20 in the number of individuals entered (per study arm).

This perspective will review some of the principles of sample size estimation for clinical trials (Table 1), selecting issues particularly relevant to IDDM prevention trials. Other sources give a comprehensive overview of the role of these factors (3–8) and convenient tabulations of the required sample size (9–11).

**TYPE OF OUTCOME**
The diagnosis of diabetes is the most straightforward end point for assessing the effectiveness of an intervention. Other types of outcome—including the loss of FPIR and the onset of autoimmunity—are possible, although less easily defined. Secondary outcomes such as these may avoid the need to wait for an extended period before the onset of diabetes, with the potential to reduce the cost and time scale of the study.

**Diagnosis of diabetes.** A long asymptomatic period of β-cell autoimmunity, during which insulin secretory capacity is progressively lost, usually precedes the onset of IDDM. The exact timing of the onset of diabetes depends not only on the rate of β-cell destruction but also on the prevailing insulin requirements, which may be altered by puberty and intercurrent infection among other factors. The onset of diabetes, defined according to WHO criteria (12), is followed by a variable (and sometimes significant) period before the diagnosis is made. In some cases the diagnosis is delayed until the onset of symptoms with unequivocal hyperglycemia, whereas in others it may occur earlier, after the discovery of asymptomatic hyperglycemia. The latter may occur more frequently among first-degree relatives of individuals with IDDM, especially if they are study participants. Trials using the diagnosis of diabetes as the outcome measure should therefore stan-
TABLE 1
Factors defining the appropriate sample size for a clinical trial

<table>
<thead>
<tr>
<th>Factors</th>
<th>Sample Size Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of outcome (binary versus continuous)</td>
<td></td>
</tr>
<tr>
<td>Desired minimal detectable effect of treatment</td>
<td></td>
</tr>
<tr>
<td>Entry criteria (homogeneity of the participants with regard to the risk and the response to intervention)</td>
<td></td>
</tr>
<tr>
<td>Degree of stratification for baseline risk factors</td>
<td></td>
</tr>
<tr>
<td>Conventional versus cost-saving trial with reduced data collection</td>
<td></td>
</tr>
<tr>
<td>Fixed versus sequential design</td>
<td></td>
</tr>
<tr>
<td>Desired power</td>
<td></td>
</tr>
<tr>
<td>Desired type I error protection</td>
<td></td>
</tr>
<tr>
<td>Number of arms and the allocation ratio</td>
<td></td>
</tr>
<tr>
<td>One-sided or two-sided alternative hypothesis</td>
<td></td>
</tr>
<tr>
<td>Anticipated distribution of entry (simultaneous versus staggered)</td>
<td></td>
</tr>
<tr>
<td>Anticipated length of patient follow-up</td>
<td></td>
</tr>
<tr>
<td>Participant drop-out rate</td>
<td></td>
</tr>
<tr>
<td>Anticipated degree of noncompliance</td>
<td></td>
</tr>
<tr>
<td>Number of comparisons</td>
<td></td>
</tr>
<tr>
<td>Number of interim looks (for safety monitoring)</td>
<td></td>
</tr>
<tr>
<td>Number of secondary outcomes</td>
<td></td>
</tr>
<tr>
<td>Anticipated treatment effect lag time</td>
<td></td>
</tr>
<tr>
<td>Direction and strength of interaction between an intervention</td>
<td></td>
</tr>
<tr>
<td>and a characteristic of participants or between interventions</td>
<td></td>
</tr>
<tr>
<td>Type of data analysis (e.g., comparison of proportions versus time to event; underlying survival distribution)</td>
<td></td>
</tr>
</tbody>
</table>

Adapted from Meinert CL and Tonascia S (3).

standardize this end point to the extent possible by performing regular OGTTS.

**Time to diabetes.** Some interventions may have the ability to delay rather than prevent the onset of IDDM. In this situation, standardizing the criteria for the diagnosis of diabetes will allow a more accurate comparison of the time taken to develop diabetes in the intervention and placebo groups. A delay in the onset of diabetes may still be considered a clinically relevant outcome if it allows the deferral of insulin dependence until after adolescence or the postponement of future diabetes complications.

**Loss of FPIR.** The use of a continuous rather than a binary outcome offers the theoretical possibility of a gain in statistical power. Several quantitative measures of pancreatic β-cell function and mass have been proposed (13). The FPIR (14,15), i.e., the sum of the insulin concentrations at 1 and 3 min during a standard IVGTT, has gained the most recognition. Unfortunately, the FPIR is cumbersome to measure, especially in young children, although alternative metabolic indexes, based on a single capillary blood sample (D.K. McCulloch, unpublished observations), may soon become available. The FPIR also is associated with a large measurement error and within-subject variation (16,17). An analysis of data from the Joslin Family Study suggests that the FPIR is probably not a feasible trial outcome. The mean change in the FPIR of ICA+ relatives during follow-up has a very large SD. Consequently, a very large sample size would be required to detect a difference between the intervention and placebo groups in a trial. In addition, it is difficult to include subjects who develop diabetes before the trial end point, as the FPIR can no longer be measured in these individuals. Thus, it is likely that clinically relevant binary end points (the diagnosis of diabetes or the need for insulin treatment) will remain the main outcome in IDDM prevention trials, rather than a surrogate metabolic outcome.

**Onset of autoimmunity.** Primary β-cell autoimmunity prevention trials are based on the premise that the autoimmune process is initiated by an environmental factor in genetically susceptible individuals. Their aim is to prevent the initiation of autoimmunity by avoiding exposure to an environmental trigger, rather than merely suppressing an established autoimmune state. To our knowledge, only one Finnish study is planning to test such an intervention, but more trials are likely to begin soon and adequate measures of their effectiveness need to be established.

Autoimmunity, rather than diabetes, might be the primary end point here, and it could be detected by using cellular or humoral assays. Humoral markers will probably become the foremost outcome measure for the primary prevention of autoimmunity. The current ICA assay by immunohistochemistry is labor intensive and difficult to standardize, despite five international standardization workshops. The cloning of β-cell autoantigens and their availability in pure form may allow the development of sensitive and reproducible assays, which are more suitable for mass screening. Besides the available radioimmunoassays for IAA+ antibodies (18,19) and antibodies to GAD (20–22), we are developing a radioassay for autoantibodies to a newly cloned islet protein, ICA69 (23). The IAA, anti-GAD, and ICA69 antibodies are each present in ~60% of prediabetic individuals and appear to occur independently from one another. Preliminary data suggest that at least one of these three is present in ~90% of susceptible relatives, and the presence of more than one greatly increases the future risk of IDDM. It is unclear whether the islet autoantibodies appear simultaneously, at random or in a fixed sequence. If one antibody precedes the others, the antigen against which it is directed may be of particular etiological importance. Such an early marker may be an especially useful outcome measure in autoimmunity prevention trials.

Autoantibody levels could either be dichotomized (positive/negative) or treated as a continuous outcome. For instance, it appears that if the IAA level in a first-degree relative is elevated above 39 nU/ml on two occasions, it remains elevated indefinitely. In the presence of one or more additional autoantibodies, such a person is very likely to develop IDDM (19). This is consistent with an all-or-none process, but how well does the 39 nU/ml cut-off point (the mean + 3SD in a control group) discriminate individuals with and without β-cell autoimmunity? To answer this question it will be necessary to define the sensitivity, specificity, and predictive values of the IAA and other quantitative autoantibody assays for the diagnosis of autoimmunity. High reproducibility within and between laboratories is needed before cut-off points are proposed. It is probable that the cut-off points will differ with age and that the tests will have lower efficiency in the general population than in relatives. In addition, some measurements may be invalid in the presence of a specific intervention. For example, IAA levels may be spuriously elevated in subjects given insulin as part of a trial.
Primary prevention of autoimmunity may eventually prove to be the only effective prevention strategy. However, prevention strategies are hampered by a lack of knowledge about the natural history of the pathogenic process leading to IDDM. Some of the critical unanswered questions relate to the apparent heterogeneity of the disease, the exact genetic basis for susceptibility, the environmental triggers, the timing of the initial event (perhaps in utero, infancy, or early childhood), the existence or otherwise of natural remissions, and the role of factors that promote or prevent progression to clinical diabetes. Before these issues are resolved it is reasonable to make conservative assumptions when calculating the sample size required for primary prevention trials. Specifically, one should anticipate that the intervention tested will have a small preventive effect or one limited only to a subset of participants.

PILOT TRIALS
Least detectable effect of the intervention. The choice of a desired detectable effect of the intervention is critical in sample size estimation; the smaller the effect, the larger the sample required to prove it. Unfortunately, many clinical trials have inconclusive results because of unrealistic expectations about the likely effect of treatment. Although a reduction in incidence as large as 50% is rarely likely, a reduction as small as 10% may nevertheless have important public health benefits. The effect of a new intervention can be estimated more accurately if pilot data regarding the new intervention are available and the incidence of the outcome in untreated subjects is known.

Nicotinamide. A pilot study to evaluate the effect of nicotinamide in ICA+ first-degree relatives has been conducted jointly at the Auckland Children's Hospital and the Barbara Davis Center in Denver (24). Currently, with a maximum of 4 yr follow-up of the subjects receiving nicotinamide, 3 of 4 Denver subjects and 4 of 10 Auckland subjects (R.B. Elliot, unpublished observations) have progressed to overt diabetes. In a study conducted at the Joslin Diabetes Center, 3 of 3 relatives treated with nicotinamide developed diabetes, and their rate of progression could not be distinguished from that of historical control subjects (25). In both studies, ICA+ relatives with low FPIR were studied, suggesting that intervention with oral nicotinamide may be ineffective in individuals with measurably impaired FPIR. Such a possibility is supported by the failure of nicotinamide, in two randomized placebo-controlled trials, to preserve β-cell function in children and adolescents with newly diagnosed IDDM (26,27), although a beneficial effect was reported in another study using a higher dose of nicotinamide in older patients (28).

Insulin. A pilot trial of intravenous and subcutaneous insulin therapy in high-risk first-degree relatives is ongoing at the Joslin Diabetes Center (2). Twelve relatives with high titer ICA, predicted time to diabetes of <3 yr (25,29) and normal oral glucose tolerances were offered participation in the trial, and 5 accepted therapy. Currently, with a maximum of 4 yr of follow-up, all 7 relatives who declined participation are now diabetic versus 1 of 5 who received insulin (P = 0.005, Fisher's exact test). The outcome of this pilot trial is by no means definitive but has prompted a larger randomized insulin trial. These preliminary data, with all their limitations, suggest that a nicotinamide trial will require a much larger sample size than an insulin trial because the preventive effect of nicotinamide, if any, appears to be smaller.

Participation rates. Participation and drop-out rates often cannot be accurately predicted in advance. The acceptance of the trial may depend on publicity, the existence of encouraging pilot data, endorsement by physicians, and other factors. An orally administered agent, such as nicotinamide, is likely to have a greater acceptance than an injected one. The low participation rate in the insulin pilot trial may have resulted from the fear of severe hypoglycemia. However, severe hypoglycemia was subsequently not observed during the pilot study, and the participation rate for relatives offered the randomized trial has been 11 of 11 to date.

ENTRY CRITERIA
Only 10% of new cases of IDDM have an affected first-degree relative. Consequently, future efforts to prevent IDDM will need to target the general population, rather than relatives, if they are to have a major impact on the incidence of the disorder. However, trials in the general population are logistically demanding because of the enormous numbers that need to be screened to identify an at-risk cohort. In addition, good long-term follow-up data on IDDM risk are available only for first-degree relatives of IDDM patients (19,30-32), and the positive predictive value of screening tests will necessarily be lower in the general population than in relatives. One approach in the general population is to use an initial screening for genetic markers followed by autoantibody testing in a genetically susceptible subset of the population. Nevertheless, nondiabetic relatives with high ICA levels currently offer a more practical and efficient avenue for the initial testing of a potential intervention. Such ICA+ relatives may be further subdivided according to several prognostic factors.

FPIR. Figure 1 illustrates the follow-up of 79 ICA+ relatives (>40 JDF U), subdivided by their FPIR on the earliest IVGTT performed. These subjects, identified at the Joslin Diabetes Center by screening nearly 10,000 first-degree relatives, consented to IVGTT before overt diabetes. The figure updates and confirms previously reported data (32) with a larger number of relatives. Relatives with FPIR <290 pM (the first percentile of the FPIR distribution in normal adults) have a significantly higher risk of IDDM. Clearly, the FPIR is an additive prognostic factor in an already highly selected population of individuals with high ICA titers.

Dual parameter model. Several other immunological (19,30,33), genetic (34), and demographic (30) factors have been reported to influence the risk of progression from β-cell autoimmunity to IDDM. A combination of these factors must have a higher predictive value than any of the factors alone. Vardi et al. (29) proposed a
predictive formula, in which the time to diabetes was modeled as a linear function of the FPIR and IAA levels in 13 ICA+ relatives (25). In Fig. 2, the same 79 ICA+ relatives as in Fig. 1 are subdivided by their predicted time to diabetes (<3 vs. ≥3 yr) from this dual parameter model. A comparison of the survival curves in Figs. 1 and 2 suggests that the dual parameter model correctly identifies more individuals at immediate risk for diabetes than the FPIR alone. At the initial IVGTT only ~30% of the ICA+ relatives had FPIR <1st percentile, whereas >50% had a predicted time to diabetes of <3 yr. It should be recognized, however, that the dual parameter model was derived from a relatively small number of individuals followed to diabetes and needs to be tested on other data sets.

Subsets of ICA. It is now clear that the autoantibodies detected with immunohistochemical ICA assays are heterogeneous. Based on different staining patterns, at least two major subsets have been identified, restricted and nonrestricted ICA (33). The restricted ICA pattern is attributable to autoantibodies to GAD (35) and is rarely associated with progression to overt diabetes in either first-degree relatives (33) or patients with polyendocrine autoimmunity (36). We estimate that ~10% of ICA+ relatives have this restricted pattern. This subgroup frequently has very high titer ICA (>640 JDF U), nearly always lacks other autoantibodies (such as IAA) and has normal FPIR. The dual parameter model correctly predicts these relatives to be at low risk.

Which of these prognostic factors should be used as entry criteria in an IDDM prevention trial? Restricting enrollment to a high-risk population, for example, individuals with low FPIR, allows for a smaller sample size because the estimated effect of the intervention is less affected by variability in the study population (3,4). Unfortunately, it usually increases the cost and time required for participant recruitment and limits the ability to make inferences about the population at large. Stratification of the trial participants by prognostic factors such as the FPIR may be a way to avoid these limitations.

STRATIFICATION FOR BASELINE RISK FACTORS
Stratification of participants may take place at two distinct phases of a clinical trial: during randomization or during analysis (poststratification). Poststratification may help identify subgroups of the study population most likely to benefit from, or be harmed by, the intervention (4) and suggest mechanisms by which a new intervention prevents disease.

Stratified randomization assures similarity of the intervention and control groups with regard to known or suspected prognostic factors. However, it rarely improves statistical efficiency for large studies (with at least 50 participants in each arm). The number of stratification variables should not exceed 2 or 3, to avoid logistical problems with balanced filling of the strata and a dramatic increase in the required sample size (5). Only variables that can be measured with high validity and reliability should be used.

Sometimes there is a priori evidence for a differential effect of the intervention by strata. For example, nicotinamide is likely to be less effective in high-risk relatives with low FPIR. If this is the case, it may be cost-efficient to randomize to each stratum only the minimum number of participants required to test the subgroup hypothesis, as illustrated in the examples below.

EXAMPLES
In the following examples, we will consider only conventional trials with fixed design and a binary outcome. The
Among 45,000 relatives screened, 900 will have high titers of ICA; 300 of these will have low FPIR and a 95% risk of IDDM over the next 4 yr, whereas 600 will have FPIR ≥290 pM and a 20% risk. With a 1:1 allocation ratio, 150 high-risk and 300 low-risk subjects will be randomized to the nicotinamide group, and the remaining 450 will be given placebo. The lowest 2 × 2 table summarizes the results of this trial without considering the initial FPIR. In the nicotinamide and control groups, respectively, 183 and 203 subjects would develop diabetes. This difference is insufficient to claim that nicotinamide is effective (with Yate’s correction $\chi^2 = 1.64; P = 0.20$), despite the huge number of relatives screened. For this unstratified analysis to yield a significant result, nicotinamide must have at least a 50% protective effect in the low-risk group (i.e., reduce the 4-yr risk of IDDM from 20 to 10%). Alternatively, the subjects must be followed for a longer time period, although this is based on increasingly less precise estimates of survival in the placebo group (Fig. 1). Considering the low-risk group separately, the 33% protection would be just discernible ($\chi^2 = 4.33; P = 0.04$). However, if any of our optimistic assumptions (see above) is in error, then the substantial protective effect in this subgroup might be missed. The failure or success of such a study may have important public health implications. If autoimmunity could be detected in the relatives before the FPIR decreases <290 pM, nicotinamide treatment could prevent (or at least delay) IDDM in 33% of them.

It may not be immediately obvious that testing a new intervention is much more cost-efficient, and should therefore be tried first, in high-risk individuals. To detect a 25% risk reduction, from 20 to 15%, 988 low-risk subjects (FPIR ≥290 pM) are needed in each arm. However, the same 25% risk reduction from 95 to 72% can be detected with only 42 high-risk subjects (FPIR <290 pM) in each arm (9).

Example 2. Although our preliminary data suggest that nicotinamide is not effective in relatives with a low FPIR, what if it is? Another study, illustrated in Fig. 4, tests the hypothesis that nicotinamide reduces IDDM risk in relatives with low FPIR (<290 pM) by at least 33%. A power calculation indicates that ~28 subjects are needed in each arm (9). With the same prevalence of high titer ICA and low FPIR as in example 1, 8400 relatives have to be screened to find 56 high-risk subjects. Over 4 yr, 27 of 28 (95%) members of the control group will develop IDDM and only 18 of 28 will become diabetic in the nicotinamide group. This difference is highly significant ($\chi^2 = 7.24; P = 0.007$, Fisher’s exact test $P = 0.003$).

The responsible use of predictive parameters may...
diminish some of the ethical problems posed by these trials (38). What besides time and money is lost in a trial designed to be large enough to detect a protective effect even if the drug works only in a subgroup of the participants? In the first example (Fig. 3), 56 high-risk relatives would suffice to detect a 33% reduction in IDDM incidence but 300 of them would be randomized. Thus, 244 more high-risk volunteers than required would be randomized. On the other hand, an important protective effect of the drug may be missed in the low-risk relatives, because of insufficient numbers in this group, because most would never develop IDDM anyway. The latter pitfall might be avoided if additional risk markers were considered and randomization limited to subjects at high risk of IDDM.

CONCLUSIONS
Most of the factors that influence the required sample size of a prevention trial can only be estimated before the trial begins. This may be especially true in IDDM prevention, for which the amount of prior experience is limited. It is advisable to protect trials against inconclusive results by making conservative assumptions, guided by pilot data.

We are excited about the prospects for finding a safe and effective intervention for preventing IDDM. However, prevention should not be tested outside the context of ethically approved research, preferably randomized double-blind, placebo-controlled trials (14). It may not be possible for trials that are being initiated today to incorporate the newest risk markers in their design, but it is imperative to refine these tools and achieve a better understanding of the natural history of prediabetic autoimmunity. Finally, collaboration among research groups, through the exchange of study protocols, standardization of measurements for future meta-analyses, and sharing of safety monitoring data, will allow optimal progress toward the goal of preventing IDDM.

ACKNOWLEDGMENTS
These studies were supported by the Blum-Kovler Foundation, the Juvenile Diabetes Foundation (Grant 1911375), the American Diabetes Association, the National Institutes of Health (Grants DK-43279, DK-32083, and DK-32493), the Joslin Diabetes Foundation Diabetes and Endocrinology Research Center Clinical Research Core (Grant IP30 DK-36837), and the Children’s Diabetes Foundation of Denver.

REFERENCES
31. Bonifacio E, Bingley PJ, Shattock M, Dean BM, Dunger D, Gale EAM, Bottoza GF: Quantitative islet cell antibody measurement