For debate

The selection of control subjects for case/control analysis of susceptibility to Type 1 (insulin-dependent) diabetes mellitus

P. Pozzilli¹, R. Buzzetti¹, G. F. Bottazzo², R. Tosi³

¹ Istituto II Clinica Medica, University of Rome "La Sapienza", Rome, Italy

² Department of Immunology, The London Hospital Medical College, London, UK

³ CNR Dipartimento Biologia Cellulare, Rome, Italy

An increasingly intense area of research in Type 1 (insulin-dependent) diabetes mellitus is directed at establishing the factors involved in susceptibility to the disease and to quantify their predictive value [1]. Each factor must be assessed first as to the *significance* of its association with the disease and secondly to the *degree* of association, usually expressed as relative risk. In both cases, the selection of appropriate control subjects is of vital importance. Unfortunately, no one type of control group is suitable for all studies and no universal criteria for what constitutes an acceptable control group can be issued. Important considerations for the selection of appropriate control groups are the biological characteristics of the disease, including genetic predisposition, aetiologic factors and mode of transmission.

A typical example of the difficulty in selecting a control group is represented by the HLA-Type 1 diabetes association studies [2] and the suitability of blood donors as controls. This, the most commonly used control group was subjected to heavy criticism at recent scientific meetings. Let us consider if and why this criticism is justified.

In an HLA-disease association study the control group should be composed of individuals representative of the HLA phenotype frequencies of the population to which the patients belong. Since there is no evidence of a variation of HLA frequencies with age, even in octogenarians, it does not seem important that the control group matches the patient group for this parameter. Also there is no evidence that blood donors, as such, have different HLA frequencies if the population they represent is homogeneous. Clearly, the major factor that is known to affect HLA phenotype frequencies is ethnic origin. Therefore, the ethnic composition of patients and control subjects must be matched. Populations which are heterogeneous because of recent migrations may be very difficult to study in this respect because a perfect ethnic match between patients and control subjects may be difficult to achieve. In such ethnically heterogeneous populations, "normal" or "healthy" blood donors may not be an appropriate choice if willingness to donate blood varies in the different ethnic groups. Thus, while in relatively homo-

geneous populations the selection of blood donors as controls is appropriate, it may not be so in other populations if willingness to donate leads to deviation in the ethnic composition of the control group. Another objection is that blood donors are selected as "healthy" individuals and that the exclusion of some pathological conditions prevents them from being representative of the general population. However, this would not seem to be a serious objection because the absence of diseases known to be HLA associated is not expected to notably alter gene frequencies due to the low penetrance of the HLA alleles involved (the only possible exception being HLA-B27 in regard to ankylosing spondylitis). The first conclusion to be drawn is that the major factor to be considered in the choice of HLA-Type 1 diabetes association studies is ethnic matching. If this is achieved, other factors such as age or health condition can be considered comparatively unimportant and blood donors may represent an appropriate control group. The problem remains, how can one achieve ethnic matching in heterogeneous populations, such as that in North America, where successive migration waves have not yet been absorbed in a common gene pool? The ethnic matching criteria should be adopted in populations which have remained comparatively isolated because of geographic and/or socioeconomical reasons and thus can be considered as ethnically homogeneous. A good example is the Sardinian population which may become very important for studies of Type 1 diabetes [3]. This relatively small population can be mapped ethnically according to their ancestral history, as demonstrated for linguistic settlements in different geographical areas following HLA analysis [4]. This has been facilitated by the low immigration rate to the island.

If we now consider other factors with reference to Type 1 diabetes susceptibility or as predictors of disease onset, the situation may be entirely different. In general, these traits will have an environmental, as well as a genetic component, and they may vary with age, sex etc. It is important that the control group is matched for the major influencing factor(s) as well as for ethnic origin. This requires considering control subjects matched for age and

sex which would be difficult using only blood donors. Most suitably, they may be schoolmates of the probands. The inclusion in the control group of individuals who have not yet developed the disease can be minimized by excluding from the control group individuals whose families have a history of Type 1 diabetes. Sporadic cases will still exist but their frequency is presumably so low as to not appreciably influence the validity of the study. In any case, they cannot be completely eliminated from the control group. To avoid this problem the siblings from diabetic families differing from the proband for two haplotypes could be selected as control subjects. The incidence of the disease in these siblings has been shown to be close to zero. However, this procedure would create another, more serious problem, i.e. the control group would be heavily selected for the *absence* of the DR3 and DR4 haplotypes. Such an uncontrolled distortion of HLA phenotype frequencies is not acceptable. This brings us to another issue. The association of certain HLA phenotypes with Type 1 diabetes has been firmly established, at least in the Caucasoid population. However, if other non-HLA factors are evaluated, how should the controls be selected with regard to their HLA representation? There are two possibilities: the first is that no selection of the control group according to HLA phenotype is made. In this case, differences detected between the disease group and the control group will reflect the sum of the influence of the factor under study plus HLA. The second possibility is to try to evaluate the influence of other factors per se minimizing HLA differences. Selecting control subjects with HLA phenotypes matching those of patients will not be easy, but still possible. A sufficient number of control subjects matching for the DR3/4 phenotype, which is represented in approximately 50 % of patients, will be almost impossible to identify. However, individuals matching for either DR3 or DR4, according to the population considered, may be found in sufficient numbers. By comparing patients and control subjects who are e.g. DR3-positive but perhaps even better matched according to DQ, one can estimate if possessing an additional factor appreciably increases the risk of the disease separate from the risk conferred by the presence of the given genetic marker per se. Such HLA matching of patients and controls may also be necessary if there is an interaction between the hypothetical additional factor and a given HLA phenotype. A good example is the stronger association offered by insulin polymorphism with Type 1 diabetes in DR4-positive diabetic patients. An effective test of this hypothesis will require typing of DR3 and DR4 Type 1

diabetic patients and comparing them to DR3 and DR4 control subjects.

In conclusion, first, the choice of appropriate control subjects cannot follow universal rules but must be dictated by the hypothesis being tested. Secondly, in HLA-association studies, the major concern should be ethnic matching of patients and control subjects. The "contamination" of disease in the control group can be minimized by eliminating individuals with family history of the disease being studied or of HLA-associated diseases in general. Thirdly, since the association with certain HLA haplotypes is assessed with certainty, at least in Caucasoid populations, this poses the problem of selection of HLA-matched control subjects in the study of other potential intervening factors.

Identification of susceptibility to Type 1 diabetes is the aim of several ongoing studies, with the ultimate goal of intervention to prevent Type 1 diabetes in genetically and immunologically susceptible individuals [5]. The highest risk should be determined on the appropriate control subjects. Therefore when assessing the risk for Type 1 diabetes the selection of such a group needs to be given serious consideration.

References

1. Bingley P, Bonifacio E, Gale E (1993) Can we really predict IDDM? *Diabetes* 42: 213–220
2. Thorsby E, Rønningen KS (1993) Particular HLA-DQ molecules play a dominant role in determining susceptibility or resistance to type 1 (insulin-dependent) diabetes mellitus. *Diabetologia* 36: 371–377
3. Songini M, Loche M, Muntoni Sa et al. (1993) Increasing prevalence of juvenile onset type 1 (insulin-dependent) diabetes mellitus in Sardinia: the military service approach. *Diabetologia* 36: 547–552
4. Cavalli Sforza LL, Minch E, Mountain JL (1992) Coevaluation of genes and languages revisited. *Proc Natl Aca Sci USA* 89: 5620–5624
5. Deschamps I, Boitard C, Hors J, Busson M, Marcelli-Barge A, Mogenet A, Robert JJ (1992) Life table analysis of the risk of type 1 (insulin-dependent) diabetes mellitus in siblings according to islet cell antibodies and HLA markers. An 8-year prospective study. *Diabetologia* 35: 951–957

Dr. P. Pozzilli
Endocrinologia (I)
Istituto II Clinica Medica
Policlinico Umberto I
I-00161 Rome
Italy