Editorial

The use of diseased control groups in genetic association studies

F. Esen, A. Celik, H. Yazici

Fehim Esen, Medical student
Ahsen Celik, Medical student
Hasan Yazici, MD

Division of Rheumatology, Cerrahpasa Medical Faculty, Istanbul University, Istanbul, Turkey.

Please address correspondence to:
Dr H. Yazici,
Cerrahpasa Tip Fakultesi Aksaray,
34098 Istanbul, Turkey.
E-mail: hyazici@attglobal.net

Received on August 28, 2009.

Clin Exp Rheumatol 2009: 27 (Suppl. 53): S4-S5.

Key words: Genetic association, specificity, polymorphism, diseased controls.

Competing interests: none declared.

In medical research, the use of diseased control groups is crucial in determining the specificity of an observation. We had noted a lack of diseased control groups in genetic association studies. This lack of specificity might also contribute to the poor reproducibility of the results of such work (1).

We formally surveyed the use of parallel diseased controls in genetic association studies in main rheumatology and internal medicine journals. We are unaware of the existence of a similar survey.

Using PubMed, we searched five prominent rheumatology (*Arthritis and Rheumatism*, *Annals of the Rheumatic Diseases*, *Journal of Rheumatology, Rheumatology*, and *Clinical and Experimental Rheumatology*) and six general medicine journals (*New England Journal of Medicine*, *Lancet, Annals of Internal Medicine, Archives of Internal Medicine, American Journal of Medicine*, and *Journal of American Medical Association*) for gene-disease association studies published before December 2007. The term "Polymorphism AND Genetic Association" was searched with the limits of "journal article", "English", "human", "added to PubMed in the last 5 years" and "published in the last 5 years". The studies selected were the ones in which the primary aim of the study was to show an association of a polymorphism with disease susceptibility. Studies in which the aim of the study was different, like associating the polymorphism with clinical outcome (complications, prognosis), drug effects etc. were excluded.

Two authors, FE and AC, independently searched through each article to survey the study designs. When a disagreement was present between the observers, FE and AC analyzed the article together (in ~10% of the time). If they still could not agree, the 3rd author (HY) made the final classification (in two manuscripts).

The articles were first classified as population-based and family-based studies. Then, the control groups of only population-based studies were checked for the presence of healthy and/or diseased controls and grouped as: (1) studies with only healthy controls; (2) studies with healthy and diseased controls; and (3) studies with only diseased controls. The family-based studies were those in which diseased and non-diseased individuals from the same family were studied. If an article included more than one type of study design in two independent studies, both of the studies were considered as separate.

Two hundred and sixty-four articles were identified; 243 (92%) were in rheumatology and 21 (8%) in general medical journals. As tabulated in Table I, 237 (89.8%) of all articles were population based studies and 27 (10.2%) were family-based. Among the population-based studies only 24 (10.1%) utilized both healthy and diseased controls and there were no appreciable differences in the frequencies of diseased control group use between the articles in the general medicine and the rheumatology journals.

Our results indicate that the specificity of the genetic association work is a particularly overlooked issue even in widely read rheumatology and general medicine journals. In around 90% of the population-based studies, the use of parallel diseased controls was not considered. We specifically focused on the population based studies in that the use of diseased controls might not be so important in the latter design.

The rather wide availability of molecular biology technology made it commonplace to resort to hypothesis-free testing of genetic variants with the hopes of finding gene-disease association. This popular inductive approach has both its proponents and opponents (2). The approach can obviously be useful as a

Table I. The study designs of the genetic association studies surveyed.

Study Design	Articles in Rheumatology journals n (%)	Articles in General Medical Journals n (%)	Total n (%)
Population-based studies	221 (91.0)	16 (76.2)	237 (89.8)
- with only healthy controls	196 (88.7)	15 (93.8)	211 (89.0)
- with only diseased controls	2 (0.9)	0	2 (0.8)
- with both healthy and diseased controls	23 (10.4)	1 (6.2)	24 (10.1)
Family-based studies	22 (9.0)	5 (23.8)	27 (10.2)
Total	243	21	264

tool to generate hypotheses later to be falsified by the time-honored deductive method. This approach, however, may be unnecessarily time consuming.

It is to be underlined here that the need for diseased controls for better specificity is not desired only for diagnostic purposes. Equally important is the need for specificity to better understand disease mechanisms in conditions of unknown etiology. The story of the description of gene mutations associated with familial Mediterrenean fever (FMF) might be a case in point. The seminal works that led to the initial description of the FMF pyrin association did not include diseased controls (3, 4). While there is no doubt this discovery helped us learn a great deal about mechanisms of inflammation in general, the initial enthusiasm to use the described associations for diagnostic purposes diminished, and rightfully, rather soon (5). We now know that there are many patients with FMF who do not carry these mutations while many patients with other diseases do. What is more, the biological importance of the pyrin association also changed meaning (6) once the specificity studies were available. We therefore suggest that had the initial FMF-pyrin work included diseased controls, for example patients with other auto or otherwise inflammatory conditions, the students of FMF would have been where they are now quite a number of years ago.

We thus propose that all genetic association studies should include diseased controls as well.

References

1. IOANNIDIS JPA: Non-replication and inconsistency in the genome-wide association setting. *Hum Hered* 2007; 64: 203-13.
2. WELCH JS, ROGLER G: Genomics and inductive reasoning: revolution, renaissance, or rhetoric? *Genet Med* 2003; 5: 476-8.
3. THE INTERNATIONAL FMF CONSORTIUM: Ancient missense mutations in a new member of the RoRet gene family are likely to cause familial Mediterranean fever. *Cell* 1997; 90: 797-807.
4. THE FRENCH FMF CONSORTIUM: A candidate gene for familial Mediterranean fever. *Nat Genet* 1997; 17: 25-31.
5. PADEH S, SHINAR Y, PRAS E *et al*.: Clinical and diagnostic value of genetic testing in 216 Israeli children with Familial Mediterranean fever. *J Rheumatol* 2003; 30: 185-90.
6. STEHLIK C, DORFLEUTNER A: COPs and POPs: modulators of inflammasome activity. *J Immunol* 2007; 179: 7993-8.