Published in final edited form as:

Pediatr Crit Care Med. 2005 May ; 6(3 0): S74–S77. doi:10.1097/01.PCC.0000161572.12676.91.

Assessing the quality of case-control association studies on the genetic basis of sepsis

Sally H. Vitali, MD and Adrienne G. Randolph, MD, MSc

Department of Anesthesia and the Division of Critical Care Medicine, Children's Hospital Boston, and the Department of Anaesthesia, Harvard Medical School, Boston, MA

Abstract

Objective—To define methodological standards that would improve the quality and reproducibility of case-control genetic association studies in sepsis.

Design—Summary of the published standards.

Results—Twelve main criteria and measures for evaluation of those criteria were developed from a variety of reviews and editorials on the subject.

Conclusions—Case-control candidate gene association studies represent an important tool for understanding the role that genes play in the pathogenesis of sepsis. It is important, however, to understand the limitations of these studies and the need for replication.

Keywords

sepsis; polymorphism; genetics; case-control studies; research design

When several different studies have evaluated the same polymorphism for associations with sepsis susceptibility or severity, their conclusions have often been conflicting. Of the ten studies evaluating the tumor necrosis factor (TNF)- α -308 G/A polymorphism for its association with mortality from sepsis, for example, four found a positive association (1–4) and six did not (5–10). These conflicts are not unique to sepsis but are prevalent among candidate gene case-control studies done for many complex, multifactorial diseases. When we seek to explain these discrepancies, it is clear that positive associations reported in one ethnic group may not be present in other ethnic groups because of difference in allele frequency (and its subsequent effect on power), and this does not negate the validity of the positive association but calls instead for appropriately powered studies in each ethnic group (11). Differences in ethnic groups can also result from the fact that even in appropriately powered studies, a gene may have differential influences in different ethnic groups. Ethnic differences, however, are not the most common reason for conflicts between study results. Epidemiologists and geneticists have noted that a lack of methodological rigor in the design and execution of many studies may account for some of the contradictory findings (11–15). Several journals have published articles advocating that more stringent criteria be applied to genetic association studies (14, 16). A few authors have proposed standards to evaluate and improve the quality of associations (11–13). Bogardus and colleagues (12) performed an analysis of the epidemiologic quality of all studies associating specific genes with disease

that were reported in four major medical journals in the year 1995. They concluded that “a substantial proportion of articles involving molecular genetic research may contain flaws in basic elements of research design and reporting,” and they pointed out the danger of developing genetic tests based on the conclusions drawn from flawed studies (12). In addition to leading to inaccurate tests, conclusions from poorly designed genetic association studies can change the direction of future studies in a way that diverts investigators from finding true associations. Therefore, we sought to assemble a set of criteria specifically for evaluation and design of genetic association studies of sepsis. Our criteria are summarized in Table 1 and explained and referenced as follows.

Choosing the Candidate Gene and Polymorphisms to Study

The first step in designing a genetic association study is to choose a gene of interest. This gene should be one whose product is biologically likely to play an important role in the pathogenesis of sepsis (11, 13). Some investigators choose to evaluate genes that have previously been studied for association with other similar diseases and conditions. Some will choose new genes for initial study based on biological mechanisms.

Once a gene of interest has been determined, normal variants in the gene must be determined by sequencing a number of individuals to determine frequency in various ethnic groups or patient populations. Polymorphisms are variations in the genetic code that are present in >1% of the population. Some of these polymorphisms are functional, having a direct effect on the function of the gene. This effect can be measured *in vitro* and/or *in vivo* (ideally both); authors should either cite the study where the function of the polymorphism was proven or provide supportive evidence of the polymorphism's effect as part of their study (11). Like the genetic association studies, studies of the functional significance of polymorphisms have also had some conflicting results; therefore, authors should seriously consider performing their own evaluation of the functional significance in the population under study (11).

The function of the majority of identified polymorphisms, however, is unknown. Many are studied because they are the initial or the only polymorphisms that have been identified in a gene to date. As genotyping methods have advanced, many important innate immunity genes have been sequenced to identify all polymorphisms. It is important to understand that individual polymorphisms within a gene have a pattern of linkage disequilibrium with other polymorphisms in the same or neighboring genes. Methods of identifying which polymorphisms to choose to best represent an entire gene have been developed (17) and are beyond the scope of this article. Because polymorphisms are not independent but are “linked” in a gene, a gene with 20 individual polymorphisms identified may have ten or fewer common (>5% in the population under study) haplotypes or combinations of polymorphisms. For example, Figure 1 shows that because of linkage disequilibrium in the TNF- α gene, there are only seven common haplotypes in individuals of European ancestry across the 13 polymorphisms in the gene. Genotyping a specific combination of polymorphisms that represent all common haplotypes helps to ensure that the entire gene is represented in the analysis. The individual polymorphisms may or may not be functional. Polymorphisms found to be associated with disease may be important themselves or may be in linkage disequilibrium with the important variant or combination of variants.

Type of Genetic Association Study

Once the gene of interest and polymorphisms to be studied are established, an investigator must decide what type of genetic association study to perform. For complex, multifactorial diseases like sepsis, the most common study design is a case-control design, in which cases are affected by sepsis and controls are not. A related study design is a cohort study where

patients presenting with a diagnosis (e.g., trauma), presenting after a specific procedure (e.g., coronary artery bypass graft), or requiring a certain level of care (e.g., intensive care unit) are followed for the development of sepsis or septic shock. Those patients who develop sepsis become the cases and those who do not become the controls. Another possible design is a family study using the biological parents of cases as controls. In this design, a transmission disequilibrium test is often employed. In the transmission disequilibrium test, a heterozygous parent is expected to transmit either allele 50% of the time. In a cohort of affected children, if the transmission of an allele is statistically different from the 50% transmission expected, then this polymorphism is significantly associated with disease susceptibility (11, 13, 18). This type of study has advantages for avoiding population stratification, which is described in more detail later. Obtaining DNA from two parents of an affected adult may be difficult, but this is more possible in pediatric studies. Because the overwhelming majority of genetic association studies in sepsis reported to date are of the case-control or cohort designs where the subgroup not developing sepsis are the controls, we have developed criteria to evaluate these studies although most of the criteria are also applicable to family-based study designs.

Reproducibility and Objectivity

Another important potential source of error in genetic association studies is genotyping error. Even standardized genotyping techniques may have a subjective component that can increase the possibility of observer bias, and observers themselves may not agree. Authors should employ two independent genotypers, and these genotypers should be blinded to the case-control status of the subject (12). In addition, it is important to repeat the genotyping in a randomly chosen subgroup of subjects using the same or a different technique to ensure genotyping accuracy (12). In a study of TNF polymorphisms and severe sepsis in trauma patients, O'Keefe and others (8) noted discrepant results from two different genotyping methods used for the TNF- α -308 locus. They used a third technique as a tiebreaker. Subsequently, the same group evaluated 291 samples for the TNF- α -308 polymorphism by different methods. They found that almost 10% of heterozygotes identified by restriction fragment length polymorphism were actually GG-homozygotes by a different sequencing method that was thought to be more accurate (19).

Choosing the Cases

One important concept of a case-control study design that has been largely ignored in the search for genes associated with sepsis is the careful selection of cases and controls. The case condition should be carefully defined. Sepsis investigators have generally used published consensus definitions (12, 13). More specific details as to the age and gender of the subjects, infecting organisms, and initial site of infection are also required to allow better understanding of the spectrum and homogeneity of the case group (12).

Choosing Controls

Sepsis is a complex syndrome that represents an interaction between a susceptible host and the right environmental trigger, a gene-environment interaction. An ideal control group, therefore, would be exposed to the relevant environmental influences but free from the case condition (11–13). Most of the published association studies in sepsis use random blood donors as a control group, and these individuals could have a genetic predisposition to sepsis but not have been exposed to the appropriate environmental triggers. Some authors avoid this problem by using a cohort study design, in which a group of patients with a predisposing condition to sepsis (e.g., trauma, pneumonia) are studied and those individuals who progress to sepsis become part of the case group. Alternatively, cases and controls can

be matched on environmental factors such as infecting organism, site of entry, or predisposing condition.

Population Stratification

Population stratification is an important concern in sepsis association study design. When cases and controls are not matched for ethnicity, confounding can result as allele frequencies may vary by ethnicity leading to artifactual associations (13, 15, 20). Perhaps the most cited example of the effect of population stratification is Knowler and colleagues' (21) study of the association of HLA haplotype and diabetes in the Pima and Papago tribes in Arizona. As shown in Table 2, their initial evaluation showed that the Gm3;5,13,14 haplotype was protective of acquiring diabetes, but a subsequent evaluation of all subjects who were full-heritage Pima or Papago found no association of the polymorphism with disease (21). Many studies of genetic associations with sepsis have matched cases and controls on ethnicity, but this becomes more difficult as studies are attempted in more diverse areas of the world. Also, race is a blurry concept, especially in the United States. One possible method of ensuring that population stratification does not exist among the cases and controls in a study is to genotype a large set of unlinked markers in all cases and controls (22). Another method of controlling for population stratification is by the use of family-based association studies and the transmission disequilibrium test, in which two parents and the case patient are genotyped to determine the rate of transmission of the candidate gene polymorphism alleles to the affected offspring (18).

Controlling for Other Confounders

Although ethnicity is a known confounder in case-control candidate gene association studies, other confounders that may play a role should be evaluated by multivariate analysis for their effect on the results of the study. In a study of associations between polymorphisms in the lipopolysaccharide binding protein and sepsis, Hubacek et al. (23) found that the presence of the less common Gly98 allele was associated with sepsis ($p < .02$) in male but not female patients. Age and gender should be analyzed, but other uncontrolled factors such as types of organisms identified, sites of infection, and other predisposing conditions should be analyzed as well.

Hardy-Weinberg Equilibrium

One way of assessing the accuracy of genotyping and/or the occurrence of significant population stratification in the control group is to ensure that the control population is in Hardy-Weinberg equilibrium (HWE) for the gene being assessed. HWE is a formula that describes the genotype frequencies in a population and can be used to track their changes from one generation to the next. If the control group of a study is not in HWE, then genotyping errors, inbreeding, genetic drift, new mutations, or population substructure has occurred (13). The closeness of the control group to HWE can be described by a chi-square goodness-of-fit test. Xu and others (24) assessed frequencies of 133 SNPs studied in 75 articles for HWE in the control population. They found that 12% of SNPs were not in HWE in the control population and attributed the problem predominately to genotyping error (24).

Power to Detect Positive Associations

Another important factor in designing genetic association studies is that the study should be sufficiently powered to detect positive associations. The number of cases required will depend on the frequency of the less common allele (11). A power calculation should be performed as part of the study design, because some very rare polymorphisms may require a very large data set to find positive associations.

Multiple Comparisons

Finally, given the difficulty and expense required to collect genomic DNA from large numbers of subjects, and the ever-increasing number of genes and SNPs to be evaluated, there is a tendency to test the same banks of case and control DNA for many different genetic associations. Unfortunately, as the number of “peeks” into the same data set increases, so does the potential for type I error due to multiple comparisons. Authors should address the need for increased significance if they are making multiple comparisons with their genetic data (11, 13). A Bonferroni correction may be too stringent and may raise the threshold for significance so high that true-positive associations are ignored. Alternatively, repeating a positive association in a separate set of cases and controls to confirm the finding reduces the likelihood of a spurious association (11).

CONCLUSIONS

Despite the important limitations given here, it is clear that case-control genetic association studies provide an important mechanism for studying the genetic component of complex diseases like sepsis. Authors should seek to minimize their risk of finding spurious associations or false-negative results by careful attention to study design. The importance of publishing negative as well as positive studies has been recognized by important journals. For example, after publishing an association between α -2-macroglobulin and Alzheimer’s disease that was not subsequently replicated by three other authors, the editors of *Nature Genetics* noted that “Reporting a genetic association means taking a risk— for authors and journals” but concluded that “only when results are disseminated widely will others attempt to verify or falsify them” (14). By applying tough standards to genetic association studies in sepsis and remaining mindful of their limitations, the risk of making false conclusions will be reduced while the chance of truly enhancing understanding of the genetic basis of sepsis will be greatly enhanced. In fact, some prestigious medical journals now require that authors replicate their findings in an independent population and report those results in the manuscript before a manuscript is considered for publication.

Acknowledgments

Supported, in part, by NIH NHLBI K23 award HL04278 (AGR) and NIH NICHD Pediatric Critical Care Scientist Development Program K12 award HD047349 (SV).

This work was supported by the Mannion Family Fund—Center for the Critically Ill Child, Division of Critical Care Medicine at Children’s Hospital Boston, the PALISI Network, and the ISF.

REFERENCES

1. Nadel S, Newport MJ, Booy R, et al. Variation in the tumor necrosis factor- α gene promoter region may be associated with death from meningococcal disease. *J Infect Dis*. 1996; 174:878–880. [PubMed: 8843235]
2. Tang GJ, Huang SL, Yien HW, et al. Tumor necrosis factor gene polymorphism and septic shock in surgical infection. *Crit Care Med*. 2000; 28:2733–2736. [PubMed: 10966243]
3. Appoloni O, Dupont E, Vandercruys M, et al. Association of tumor necrosis factor-2 allele with plasma tumor necrosis factor- α levels and mortality from septic shock. *Am J Med*. 2001; 110:486–488. [PubMed: 11331061]
4. Mira JP, Cariou A, Grall F, et al. Association of TNF2, a TNF- α promoter polymorphism, with septic shock susceptibility and mortality: A multicenter study. *JAMA*. 1999; 282:561–568. [PubMed: 10450718]
5. Reid CL, Perrey C, Pravica V, et al. Genetic variation in proinflammatory and antiinflammatory cytokine production in multiple organ dysfunction syndrome. *Crit Care Med*. 2002; 30:2216–2221. [PubMed: 12394947]

6. Read RC, Camp NJ, di Giovine FS, et al. An interleukin-1 genotype is associated with fatal outcome of meningococcal disease. *J Infect Dis.* 2000; 182:1557–1560. [PubMed: 11023482]
7. Schaaf BM, Boehmke F, Esnaashari H, et al. Pneumococcal septic shock is associated with the interleukin-10-1082 gene promoter polymorphism. *Am J Respir Crit Care Med.* 2003; 168:476–480. [PubMed: 12746253]
8. O'Keefe GE, Hybki DL, Munford RS. The G→A single nucleotide polymorphism at the –308 position in the tumor necrosis factor- α promoter increases the risk for severe sepsis after trauma. *J Trauma.* 2002; 52:817–825. [PubMed: 11988644]
9. Waterer GW, Quasney MW, Cantor RM, et al. Septic shock and respiratory failure in community-acquired pneumonia have different TNF polymorphism associations. *Am J Respir Crit Care Med.* 2001; 163:1599–1604. [PubMed: 11401880]
10. Stuber F, Petersen M, Bokelmann F, et al. A genomic polymorphism within the tumor necrosis factor locus influences plasma tumor necrosis factor- α concentrations and outcome of patients with severe sepsis. *Crit Care Med.* 1996; 24:381–384. [PubMed: 8625623]
11. Daly AK, Day CP. Candidate gene case-control association studies: Advantages and potential pitfalls. *Br J Clin Pharmacol.* 2001; 52:489–499. [PubMed: 11736857]
12. Bogardus ST Jr, Concato J, Feinstein AR. Clinical epidemiological quality in molecular genetic research: The need for methodological standards. *JAMA.* 1999; 281:1919–1926. [PubMed: 10349896]
13. Silverman EK, Palmer LJ. Case-control association studies for the genetics of complex respiratory diseases. *Am J Respir Cell Mol Biol.* 2000; 22:645–648. [PubMed: 10837359]
14. Freely associating. *Nat Genet.* 1999; 22:1–2. [PubMed: 10319845]
15. Risch NJ. Searching for genetic determinants in the new millennium. *Nature.* 2000; 405:847–856. [PubMed: 10866211]
16. Crossman D, Watkins H. Jestling Pilate, genetic case-control association studies, and Heart. *Heart.* 2004; 90:831–832. [PubMed: 15253943]
17. Sebastiani P, Lazarus R, Weiss ST, et al. Minimal haplotype tagging. *Proc Natl Acad Sci U S A.* 2003; 100:9900–9905. [PubMed: 12900503]
18. 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.* 1993; 52:506–516. [PubMed: 8447318]
19. Peters DL, Barber RC, Flood EM, et al. Methodologic quality and genotyping reproducibility in studies of tumor necrosis factor –308 G→A single nucleotide polymorphism and bacterial sepsis: Implications for studies of complex traits. *Crit Care Med.* 2003; 31:1691–1696. [PubMed: 12794406]
20. Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *Lancet.* 2003; 361:598–604. [PubMed: 12598158]
21. Knowler WC, Williams RC, Pettitt DJ, et al. Gm3;5,13,14 and type 2 diabetes mellitus: An association in American Indians with genetic admixture. *Am J Hum Genet.* 1988; 43:520–526. [PubMed: 3177389]
22. Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet.* 1999; 65:220–228. [PubMed: 10364535]
23. Hubacek JA, Stuber F, Frohlich D, et al. Gene variants of the bactericidal/permeability increasing protein and lipopolysaccharide binding protein in sepsis patients: Gender-specific genetic predisposition to sepsis. *Crit Care Med.* 2001; 29:557–561. [PubMed: 11373419]
24. Xu J, Turner A, Little J, et al. Positive results in association studies are associated with departure from Hardy-Weinberg equilibrium: Hint for genotyping error? *Hum Genet.* 2002; 111:573–574. [PubMed: 12516594]

	282	352	1009	1078	1440	1893	2642	4013	4040	4101	4366	4371	4765	Haplo	Freq
G	G	G	G	A	A	A	G	T	G	G	T	G		11	0.24
A	G	G	G	A	A	A	G	T	A	G	T	G		5	0.11
G	A	A	G	G	G	A	G	C	G	A	C	G		3	0.07
G	G	G	A	A	A	A	G	T	G	G	T	T		4	0.09
G	G	G	G	A	A	A	G	T	G	G	C	G		7	0.15
G	G	G	G	A	A	A	G	T	G	G	T	G		7	0.15
G	G	G	G	A	A	A	G	T	G	G	C	G		2	0.04

Figure 1.

Examples of common haplotypes of 13 polymorphisms across the tumor necrosis factor- α gene in white people. Seven haplotypes occur in 85% of the white population genotyped.

The color coding shows linkage disequilibrium across common polymorphisms in the gene; SNPs in high linkage disequilibrium are the same color. SNP numbering from <http://pga.gs.washington.edu/data/tnf/>

Table 1

Criteria for quality evaluation of case-control candidate gene association studies in sepsis

Criteria	Evaluation
1. Biological reasonability (11, 13)	Biologically functional or plausible effect of the gene product in sepsis
2. Choice of polymorphisms (11)	For functional polymorphism, evidence of function presented OR For nonfunctional SNPs, gene completely represented with test of linkage disequilibrium and haplotype analysis performed
3. Reproducibility (12)	Quality control methods used to ensure genotyping accuracy (repetition, two methods, two genotypers, etc.)
4. Objectivity (12)	Lab technicians blinded to phenotype and clinical investigators blinded to genotype
5. Sepsis definition (12, 13)	Criteria for sepsis affection clearly define cases from controls
6. Representativeness of cases (12)	Spectrum of disease represented by cases reported (severity of illness, operative status, organisms identified, sites of infection) Number of eligible patients excluded from study and reasons described
7. Case-control environment similarity (11–13)	Cases and controls matched on environmental risk factors for sepsis
8. Population stratification (13, 15, 20)	Cases and controls matched for ethnicity AND/OR Multiple unlinked markers tested in cases and controls
9. Control for confounders (13)	A table with possible confounding variables present (age, gender, types of organisms, operative status, sites of infection, etc.) A statistic for multivariate analysis for each potential confounder
10. Equilibrium of polymorphism in controls (13)	Calculate and report Hardy-Weinberg equilibrium of control population with statistical test
11. Power to detect positive associations (11)	Calculation showing how number of cases and controls provides sufficient power
12. Correction for multiple comparisons (11, 13)	Correction for multiple comparisons if case or control DNA used to test for more than one association in this study or in previous studies AND/OR Study repeated in an independent group of cases and controls

Table 2

Example of population stratification

	Total Subjects	% Diabetes
General tribe members		
Gm+ allele		
Present	293	8
Absent	4,627	29
Full-heritage Pima-Papago adults		
Gm+ allele		
Present	17	59
Absent	1,764	60

The probability of diabetes is higher in the general population of Pima and Papago Indians from southern Arizona in people with absence of the Gm+ allele. The probability is the same, however, when the analysis is restricted to full-heritage adult Pima and Papago Indians, who are less likely to have the allele but more likely to have diabetes whether or not they have the allele. From Ref. 21.