MINIREVIEW Genomics & Informatics Vol. 4(2) 51-56, June 2006

Sample Size and Power Estimation in Case-Control Genetic Association Studies

Chul Ahn

Department of Medicine, University of Texas Medical School, Houston, TX 77030, USA

Abstract

In planning a genetic association study, it is necessary to determine the number of samples to be collected for the study in order to achieve sufficient power to detect the hypothesized effect. The case-control design is increasingly used for genetic association studies due to the simplicity of its design. We review the methods for the sample size and power calculations in case-control genetic association studies between a marker locus and a disease phenotype.

Keywords: case-control study, sample size, genetic association

Introduction

Genetic association studies assess correlations between a disease phenotype and a marker genotype on a population or family-based scale. In some respects, testing for an association between a disease phenotype and a marker locus is not different from testing for statistical association between any two variables, regardless of being genetic or not. Thus, many statistical methods used to analyze ordinary statistical association are also used for genetic association. However, there are important differences between genetic association and statistical association in the study design and analysis.

Genetic association studies have been mostly performed in a case-control study setting that has been well developed for use in epidemiological methods (Romero *et al.*, 2002). The case-control design is increasingly used for genetic association studies due to its simplicity in design. Significant differences in allele frequencies or genotype frequencies between cases and controls are taken as evidence for involvement of an allele or genotype in disease susceptibility. In planning a genetic association study, it is necessary to determine the number of samples to be collected for the study in order to achieve sufficient power to detect the hypothesized effect. A genetic association study that is conducted without attention to sample size or power information takes the risks of either failing to detect important meaningful differences (i.e., type II error) or taking an unnecessarily excessive number of samples for a study. Either case fails to adhere to the Ethical Guidelines of the American Statistical Association which says "Avoid the use of excessive of inadequate number of research subjects by making informed recommendations for study size" (American Statistical Association, 1999). While the exact final number that will be used for the analysis will be unknown due to missing information such as lack of genotype information and clinical information, it is still desirable to determine a target sample size based on the proposed study design.

The sample size estimate will allow the estimation of total cost of the proposed study required. Typically, the number of samples is computed to provide a fixed level of power under a specified alternative hypothesis. The alternative hypothesis usually represents a minimal meaningful difference in the frequency of genotypes or alleles between cases and controls. Power (1-probability of type II error) is an important consideration for several reasons. Low power can cause a true difference in allele frequencies or genotype frequencies between cases and controls to be rejected. However, too much power may make results statistically significant that are not meaningfully different. The probability of type I error (α) of 0.05 and power of 0.80 and 0.90 has been widely used for the sample size estimation in genetic association studies. In this paper, we restrict our attention to the sample size and power calculations for case-control genetic association studies between a marker locus and a disease phenotype on a population- based scale.

Sample size determination based on allelic frequencies

Either the candidate allelic or genotypic frequencies are compared to evaluate genetic associations between cases and controls. The geneticist usually prefers comparing allelic frequencies on the grounds that, in the absence of an association, sampling genotypes under Hardy-Weinberg Equilibrium (HWE) is equivalent to

*Corresponding author: E-mail Chul.W.Ahn@uth.tmc.edu, Tel +1-713-500-6701, Fax +1-713-500-6722
Accepted 31 May 2006

sampling gametes, or alleles from the gene pool that gives rise to the population (Lalouel and Rohrwasser, 2002). When allele frequencies are compared, the two alleles of each subject are treated as independent. Even though comparison of allele frequencies seems to be counter-intuitive because alleles sometimes do not act independently, this approach has been often used in genetic association studies (e.g. Apple *et al.*, 1994; Odunsi *et al.*, 1995). Power and sample size for chi-square tests for comparing allelic frequencies can be found in Lachin (1977) and Guenther (1973). There are many available statistical software programs that can be used to estimate the sample size or power needed to detect minimal allele frequencies between cases and controls. For example, commercial software such as nQuery (Elashoff, 2005) or Power and Precision (Borenstein *et al.*, 2001) or StatXact (Mehta and Patel, 2006) or non-commercial software such as PS (Power and Sample size calculation, which is available at http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/PowerSampleSize) can be used to estimate the sample size or power based on allele frequencies of cases and controls.

Sample size determination based on genotype frequencies

There are two commonly used statistical tests for the estimation of sample size and power using genotypes in genetic association studies. One is the Cochran-Armitage linear trend test (Armitage, 1955) and the other is a genotypic association test (Gordon *et al.*, 2002).

1. Cochran-Armitage Linear Trend Test

Sasieni (1997) investigates the relationship between the approaches of using alleles and genotypes, and concludes that the allelic chi-square test based on alleles is inappropriate when genotype frequencies violate HWE and that the Cochran-Armitage's linear trend test should be used instead. The two tests are asymptotically equivalent under HWE. Kang *et al.*, (2005) investigate performances of the allelic chi-square test when HWE is violated. They investigate the maximum and minimum values of type I error rate of the allelic chi-square test by using nonlinear programming and simulation. It is revealed that, under violated HWE, the allelic chi-square test has distorted type I error rate. Allele frequencies should be compared only when genotype frequencies are in HWE.

In a single nucleotide polymorphism (SNP) with two possible alleles, 'A' or 'B', there are three possible genotypes, 'AA', 'AB', and 'BB'. Study subjects can be classified into a 2×3 table based on each subject's disease status (case or control) and genotype. Table 1 shows the genotype data from case-control studies. Here, the two rows refer to disease status (cases and controls), and the 3 columns refer to the genotypes.

Table 1. Data from case-control studies

	AA	AB	BB	Total
Cases	n_{10}	n_{11}	n_{12}	N_1
Controls	n_{00}	n_{01}	n_{02}	N_0
Total	n_0	n_1	n_2	N

Given disease status, the distribution of genotypes is multinomial with parameter vectors $p=(p_0, p_1, p_2)$ for cases and $q=(q_0, q_1, q_2)$ for controls, where the subscript corresponds to the number of high risk alleles in the genotype. Slager and Schaid (2001) specify the values of p_i and q_i using the information on penetrances (f_0, f_1, f_2) and the marginal genotype probabilities (g_0, g_1, g_2). Here, the penetrance f_j is the conditional probability P(affected| j copies of disease allele), where j=0, 1, or 2 for a di-allelic disease locus. The information on penetrance and the marginal genotype probabilities can be obtained from the prior data based on the literature or pilot study or an assumed genetic model. The values of p_i and q_i can be calculated from the following equations.

$$p_i = f_i g_i / (\Sigma f_i g_i),$$
$$q_i = (1-f_i) g_i / \{\Sigma (1-f_i) g_i\}$$

Once p_i and q_i are computed from fi and gi, sample size software for Cochran-Armitage trend test can be used to estimate the sample size or power needed for the association between genotypes and disease status. Slager and Schaid (2001) provide a sample size formula for Cochran-Armitage trend test, and evaluate the performances of their sample size formula under different disease prevalence, genotypic relative risks and allelic effects such as multiplicative, additive, dominant and recessive models based on high-risk allele frequencies. Cochran-Armitage trend test can be obtained as the score test using the logistic regression with a single covariate depending on the individual's genotype. Cochran-Armitage trend test is a locally optimal test for the given set of scores. Slager and Schaid (2001) demonstrate that their sample size formula is generally quite accurate by showing that their empirical powers are close to the expected powers through simulation.

When the variance of the Cochran-Armitage trend test statistic is known, the formulas for sample size and power are readily obtained. In practice, the variance of the test statistics is unknown, and needs to be estimated. Freidlin *et al.* (2002) present formulas for the sample size

needed for case-control genetic association studies at a pre-specified power when the variance of the test statistic is unknown. Since these values are typically unknown for complex diseases such as diabetes, cancer, and cardiovascular disease, one considers a range of values for parameters to deal with such uncertainty. Cox and Hinkley (1974) suggest reporting median sample size and/or power values or worst-case scenarios that yield the largest sample size needed or the smallest power. This suggestion has been used in several genetic applications (Cousin *et al.*, 2003; Gordon *et al.*, 2005; Zheng and Tian, 2005). This suggestion leads us to observe a distribution of sample size (or power) values for the range of parameter values considered, including minimum, maximum, median, and average sample size (or power) values.

2. Genotypic association test

Association between disease status and genotype is tested using the chi-square test statistic on 2×3 contingency tables. If the chi-square test yields a significant result, then we may say that the marker locus seems to be in close proximity to such a susceptibility locus for a disease. Sample size or power can be estimated using the noncentrality parameter, which can be computed from the asymptotic distribution of the chi-square test on 2×3 contingency tables under a specified alternative hypothesis (Mitra, 1958). Sample size estimate software such as nQuery can be used to estimate the sample size or power from 2×3 contingency tables. Purcell *et al.* (2003) give description of Genetic Power Calculator (GPC) package which performs power calculations for the design of linkage and association genetic mapping studies of complex traits. GPC software can be obtained from the website http://pngu.mgh.harvard.edu/~purcell/gpc/.

Power and Sample size estimation in the presence of genotype errors

There is considerable mistyping in most genotype data due to biochemical anomalies, human oversights, or shortcomings in genotyping scoring software (Sobel *et al.*, 2002). Genotyping errors lead to inflation in type I errors or a decrease in power, and bias in statistical inference in genetic association studies.

Gordon *et al.*, (2001, 2002, 2003) quantify the genotyping error effects on the sample size and power needed to maintain asymptotic type I error and type II error rates for case-control genetic association studies between a SNP locus and a disease phenotype. Software PAWE (Power for Association With Errors) to estimate the sample size and power for genetic case-control association studies allowing for genotyping error can be found from http://linkage.rockefeller.edu/pawe/. Software PAWE provides power and sample size calculations for case-control studies with a di-allelic locus such as SNP in the presence of genotyping errors. The results obtained from data without genotyping errors by software PAWE are identical to those obtained by other genetic association test power calculators, for example, the Genetic Power Calculator (GPC). Software PAWE-3D (Gordon *et al.*, 2005) plots for power and sample size calculations over a range of parameters including genotypic errors for the Cochran-Armitage linear trend test and the genotypic association test. Software is available from the website http://linkage.rockefeller.edu/pawe3d/.

Kang *et al.*, (2004) examine which genotype misclassification errors are most costly, in terms of increased sample size to maintain asymptotic power and significance level. They use chi-square test for 2×3 contingency tables instead of 2×2 contingency tables since the allelic chi-square test based on alleles is inappropriate when genotype frequencies violate HWE. Use of allelic test may be inappropriate even in HWE when genotyping errors are present. They show that misclassifying a more common genotype as the less common genotype requires a larger sample size than the reverse misclassification.

Power and Sample size estimation in the presence of phenotype errors

Edwards *et al.* (2005) describe high phenotype misclassification rates cited in several studies. Edwards *et al.* (2005) present an impact of phenotype error, also known as diagnostic error, on power and sample size estimates for case-control genetic association studies between a disease phenotype and a marker locus. The sensitivity and specificity of the diagnostic test are defined as sensitivity=Pr(observed case | true affected) and specificity=Pr(observed control | true unaffected). Two types of diagnostic errors are: Pr(observed control | true affected) = 1 - sensitivity and Pr(observed case | true unaffected) = 1 - specificity.

An example of diagnostic errors involves the use of the prostate specific antigen (PSA) test for diagnosis of prostate cancer. Although prostate cancer is the second leading cause of cancer deaths in men and accounted for an estimated 220,900 new cases and 28,900 deaths in the United States in 2003, screening for it with the PSA blood test and digital rectal exam (DRE) remains controversial. Screening cannot predict the kind of cancer

a man may have and who will die from it. As a screening test, PSA has a sensitivity of about 67.5% to 80% and a specificity of about 60% to 70% when the level is greater than 4.0 ng/mL (American Urological Association, 2000; Catalona *et al.*, 1991; Brawer, 1999; Oesterling *et al.*, 1993; Carter *et al.*, 1992). Although autopsy studies have shown that approximately one in three men over age 50 has histological evidence of prostate cancer, most have clinically insignificant disease and only about 3% of men will die from it (American Urological Association, 2000).

Edwards *et al.* (2005) conduct a genotypic test of association using the Pearson chi-square test statistic on 2×3 contingency tables in the presence of phenotype errors. Sample size or power is obtained as a function of noncentrality parameter expressed in terms of sample sizes, genotype frequencies, and phenotype error model parameters. Zheng and Tian (2005) investigate the impact of phenotype errors on sample size and power using the Cochran-Armitage trend test under various genetic models. Both Edwards *et al.* (2005) and Zheng and Tian (2005) make a key assumption that phenotype errors are random and independent, and furthermore they are non-differential with respect to a particular genotype. Both Zheng and Tian (2005) and Edwards *et al.* (2005) illustrate that phenotype misclassification errors can substantially decrease the asymptotic power to detect association between a trait locus and a marker locus. Software PAWE-PH to estimate power or sample size under random phenotype misclassification errors or diagnostic errors is available from the website http://linkage.rockefeller.edu/pawe/paweph.htm.

One of interesting findings is that for many diseases in which prevalence is less than or equal to 10%, it is much more important to insure that cases are truly cases rather than controls being truly controls (Edwards *et al.*, 2005, Zheng and Tian, 2005). That is, it is much more important to have less phenotype errors in cases than in controls.

Power and Sample size estimation in the presence of genotype and phenotype errors

Ji *et al.* (in press) derive the sample size and power estimates using the noncentrality parameter for the Likelihood Ratio Test Allowing for Error (LRTae) in the presence of random phenotype and genotype errors. They justify the sample size and power through simulations using various combinations of genotype frequencies, phenotype and genotype misclassification probabilities, total sample size, and ratio of cases to controls. Contour plots can be generated using the software available at the website http://linkage.rockefeller.edu/lrtae/contour.html.

Examples

Afshar-Kharghan *et al.* (2004) investigate the association between the risk of coronary heart disease and the variable number of tandem repeat (VNTR) polymorphism of platelet glycoprotein Ib α (GP Ib α). It is shown that CC genotype is associated with a lower risk of CHD events than the other genotypes in African-Americans. Kang *et al.* (2004) show that HWE holds well for GP Ib α VNTR from the data of Afshar-Kharghan *et al.* (2004). The B allele frequencies of case and controls in African-Americans are 0.287 and 0.146 for cases and controls, respectively.

Suppose that we wish to estimate the sample size needed to obtain 90% power at a 5% significance level with an equal number of case and controls for the chi-square test on genotypes. The sample size calculations are conducted using the software PAWE (http://linkage.rockefeller.edu/pawe/) as follows: on the first page, 'Sample size calculations for a fixed power', 'Genetic model free method', 'Douglas Skol Boehnke', 'Significance level=0.05'; on the second page, 'Power level between 0 and 1=0.90', 'Ratio of controls to cases=1', 'Hardy Weinberg Equilibrium for cases and controls=yes', 'γ=0.02' and 'η=0.02'; on the third page, 'case 1 allele frequency=0.287' and 'control 1 allele frequency=0.146'. Then, the required sample size estimates are 111 cases and 111 controls when no errors are present (When no errors are present, error model parameters are not important since we only look at the sample size estimates from the results of 'Data Without Error'). Let's assume that the probability of misclassifying homozygote as heterozygote is 2% and the probability of misclassifying heterozygote as homozygote is 2%. It is assumed that homozygotes are not misclassified as other homozygotes. The minimal sample sizes required are 117 cases and 117 controls when random genotype errors are present. Thus, the presence of genotype errors results in a 5.94% increase in sample size to maintain the same statistical power.

Discussion

In this paper, we limit our attention to the sample size and power calculations for case-control genetic association studies between a marker locus and a disease phenotype on a population-based scale. In addition to genetic association studies using unrelated individuals on a population-based scale, there are family-based genetic

association studies, which sample families through one or more probands. Family-based genetic association studies are originally developed to deal with problems of population stratification. Sample size or power of the family-based association studies can be estimated with the family-based association test (FBAT) statistic (Laird *et al.* (2000)) for virtually any given study design and ascertainment conditions with the interactive software package PBAT (http://www.biostat.harvard,edu/~clange/default.htm).

Complex diseases such as hypertension and their complications are the consequence of interactions between many genetic and environmental factors. In general, the genes contributing to risk of hypertension do not have the same effects in different environments, or in different genetic backgrounds. Genetic effects may involve gene-gene interactions as well as gene-environment interactions. The software Quanto (http://hydra.usc.edu/gxe/) computes sample size or power for association studies of genes, gene-environment interaction, or gene-gene interaction for a variety of study designs such as matched case-control, case-sibling, case-parent, and case-only designs.

In addition to case-control association designs, cohort designs are also commonly used to test genetic association between a disease phenotype and a marker locus when the sample unit consists of unrelated individuals. Some studies such as LUMINA study (Reveille *et al.*, 1988) and GENISOS study (Reveille *et al.*, 2001) have begun to employ cohort designs to study genetic associations using unrelated individuals.

The advantages of cohort studies over case-control studies are ascertainment of cases, elimination of recall bias, and reduction of possible population stratification. The disadvantages are cost and time needed to recruit and measure genetic and environmental data over time. When the outcome (phenotype) measurements are made over time in cohort studies, sample size for testing genetic association can be estimated using generalized estimating equation (GEE) approaches for continuous and binary outcomes. (Jung and Ahn, 2003, 2005; Ahn and Jung, 2003, 2005). Jung and Ahn (2003, 2005) provide a closed-form sample size formula by accounting for the missing data and the correlation between repeated measurements.

References

Afshar-Kharghan, V., Aleksic, N., Ahn, C. Boerwinkle, E., Wu, K., and Lopez, J. (2004). The variable number of tandem repeat polymorphism of platelet glycoprotein Ib α and risk of coronary heart disease. *Blood* 103, 963-965.

Ahn, C. and Jung, S. (2003). Efficiency of GEE estimators of slopes in repeated measurements: Adding subjects or adding measurements? *Drug Information Journal* 37, 309-316.

Ahn, C. and Jung, S. (2005). Effect of dropout on sample size estimates on trends across repeated measurements. *J. of Biopharmaceutical Statistics* 15, 33-41.

American Statistical Association. (1999). Ethical guidelines for statistical practice: Executive Summary. *Amstat News April,* 12-15.

American Urological Association. (2000). Prostate-specific antigen (PSA) best practice policy. *Oncology* 14, 267-280.

Apple, R., Erlich, H., Klitz, W., Manos, M., Becker, T., and Wheeler, C. (1994). HLA DR-DQ associations with cervical carcinoma show papillomavirus-type specificity. *Nature Genetics* 6, 157-162.

Armitage, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics* 11, 375-386.

Borenstein, M., Rothstein, H., and Cohen, J. (2001). *Power and Precision* Biostat, Inc. Englewood, NJ.

Brawer, M.K. (1999). Prostate-specific antigen: current status. *CA Cancer J. Clin.* 49, 264-281.

Carter, H.B., Pearson, J.D., Metter, E.J., Brant, L.J., Chan, D.W., Andres, R., Fozard, J.L., and Walsh, P.C. (1992). Longitudinal evaluation of prostate-specific antigen levels in men with and without prostate disease. *JAMA* 267, 2215-2220.

Catalona, W.J., Smith, D.S., Ratliff, T.L., Dodds, K.M., Coplen, D.E., Yuan, J.J., Petros, J.A., and Andriole, G.L. (1991). Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. *N. Engl. J. Med.* 324, 1156-1161.

Cox, D. and Hinkley, D. (1974). *Theoretical Statistics.* (Boca Raton: Chapman and Hall/CRC Press).

Cousin, E., Genin, E., Mace, S., Ricard, S., Chansac, C., del Zompo, M., and Deleuze, J. (2003). Association studies in candidate genes: strategies to select SNPs to be tested. *Human Heredity* 56, 151-159.

Douglas, J., Skol, A., and Boehnke, M. (2002). Probability of detection of genotyping errors and mutations as inheritance inconsistencies in nuclear-family data. *American Journal of Human Genetics* 70, 487-495.

Edwards, B., Haynes, C., Levenstien, M., Finch, S., and Gordon, G. (2005). Power and sample size calculations in the presence of phenotype errors for case/control genetic association studies. *BMC Genetics* 2005, 6, 18

Elashoff, J. (2005). *nQuery Advisor® Version 6.0 User's Guide.* (Los Angeles, CA).

Freidlin, B., Zheng, G., Li, Z., and Gastwirth, J. (2002). Trend tests for case-control studies of genetic markers: power, sample size and robustness. *Human Heredity* 53, 146-152.

Gordon, D., Heath, S.C., Liu, X., and Ott, J. (2001). A transmission disequilibrium that allows for genotyping errors in the analysis of single nucleotide polymorphism data. *American Journal of Human Genetics* 69, 371-380.

Gordon, D., Finch, S., Nothnagel, M., and Ott, J. (2002). Power and sample size calculation for case-control genetic association tests when errors are present: Application to single nucleotide polymorphisms. *Human Heredity* 54, 22-33.

Gordon, D., Levenstien, M., Finch, S., and Ott, J. (2003). Errors and linkage disequilibrium interact multiplicatively when computing sample sizes for genetic case-control association studies. *Pacific Symposium on Biocomputing* 490-501.

Gordon, D., Haynes, C., Blumenfeld, J., and Finch, S. (2005). "PAWE-3D: visualizing Power for Association With Error in case/control genetic studies of complex traits", *Bioinformatics* 21, 3935-3937.

Guenther, W. (1977). Power and sample size for approximate chi-square tests. *Am Stat* 31, 83-85.

Ji, F., Yang, Y., Haynes, C., Finch, S., and Gordon, D. (in press) Computing asymptotic power and sample size for case-control genetic association studies in the presence of phenotype and/or genotype misclassification errors. *Statistical Applications in Genetics and Molecular Biology*.

Jung, S. and Ahn, C. (2003). Sample size estimation for GEE method for comparing slopes in repeated measurements data. *Statistics in Medicine* 22, 1305-1315.

Jung, S. and Ahn, C. (2005). Sample size for repeated binary measurements using GEE. *Statistics in Medicine* 24, 2583-2596.

Kang, S., Gordon, D., and Finch, S. (2004). What SNP genotyping errors are most costly for genetic association studies? *Genetic Epidemiology* 26, 132-141.

Kang, S., Finch, S., Haynes, C., and Gordon, D. (2004). Quantifying the percent increase in minimum sample size for SNP genotyping errors in genetic model-based association studies. *Human Heredity* 58, 139-144.

Kang, S., Shin, D., Oh, M., and Ahn, C. (2004). An investigation on the allelic chi-square test used in genetic association studies. *Biometrical Journal* 46, 699-706.

Lachin, J. (1977). Sample size determination for comparative trials. *Biometrics* 33, 315-324.

Laird, N., Horvath, S., and Xu, X. (2000). Implementing a unified approach to family based tests of association. *Genetic Epi*. 19, S36-S42.

Lalouel, J. and Rohrwasser, A. (2002). Power and replication in case-control studies. *American Journal of Hypertension* 15, 201-205.

Mehta, C. and Patel, N. (2006). *StatXact 7 with Cytel Studio Cytel Software Corporation*, (Cambridge, *MA*).

Mote, V.L. and Anderson, R.L. (1965). An investigation of the effect of misclassification on the properties of chi-square tests in the analysis of categorical data. *Biometrika* 52, 95-109.

Odunsi, K., Terry, G., Ho, L., Bell, J., Cuzick, J., and Ganesan, T. (1995). Association between HLA DQB1*03 and cervical intra-epithelial neoplasia. *Molecular Medicine* 1, 161-171.

Oesterling, J.E., Cooner, W.H., Jacobsen, S.J., Guess, H.A., and Lieber, M.M. (1993). Influence of patient age on the serum PSA concentration. An important clinical observation. *Urol. Clin. North Am.* 20, 671-680.

PS: Power and sample size calculation. Software webpage is http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/PowerSampleSize

Purcell, S., Cherny, S., and Sham, P. (2003). Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 19, 149-150.

Reveille, J., Moulds, J., Ahn, C., Friedman, A., Baethge, B., Roseman, J., Stratton, K., and Alarcon, G. (1988). Systemic lupus erythematosus in three ethnic groups: I. The effects of HLA class II, C4, and CR1 alleles, socioeconomic factors, and ethnicity at disease onset. LUMINA Study Group. Lupus in minority populations, nature versus nurture. *Arthritis Rheum* 41, 1161-1172.

Reveille, J., Fischbach, M., McNearney, T., Friedman, A., Aguilar, M., Lisse, J., Fritzler, M., Ahn, C., Arnett, F., and GENISOS Study Group. (2001). Systemic sclerosis in 3 US ethnic groups: a comparison of clinical, sociodemographic, serologic, and immunogenetic determinants. *Semin Arthritis Rheum* 30, 332-346.

Romero, R., Kuivaniemi, H., Tromp, G., and Olson, J. (2002). The design, execution, and interpretation of genetic association studies to decipher complex disease. *Am. J. Obstet. Gynecol.* 187, 1299-1312.

Sasieni, P. (1997). From genotypes to genes: Doubling the sample size. *Biometrics* 53, 1253-1261.

Slager, S. and Schaid, D. (2001). Case-control studies of genetic markers: power and sample size approximations for Armitage's test for trend. *Human Heredity* 52,149-153.

Sobel, E., Papp, J., and Lange, K. (2002). Detection and integration of genotyping errors in statistical genetics. *Am. J. Hum. Genet.* 70, 496-508.

Zheng, G., and Tian, X. (2005). The impact of diagnostic error on testing genetic association in case-control studies. *Statistics in Medicine* 24, 869-882.