Genetic Epidemiology 24: 291–296 (2003)

Simple Method to Analyze SNP-Based Association Studies Using DNA Pools

Peter M. Visscher[1]* and Stéphanie Le Hellard[2]

[1]*Institute of Cell, Animal, and Population Biology, University of Edinburgh, Edinburgh, United Kingdom*
[2]*Medical Genetics Section, Molecular Medicine Centre, University of Edinburgh, Edinburgh, United Kingdom*

Association studies using DNA pools are in principle powerful and efficient to detect association between a marker allele and disease status, e.g., in a case-control design. A common observation with the use of DNA pools is that the two alleles at a polymorphic SNP locus are not amplified in equal amounts in heterozygous individuals. In addition, there are pool-specific experimental errors so that there is variation in the estimates of allele frequencies from different pools that are from the same individuals. As a result of these additional sources of variation, the outcome of an experiment is an *estimated* count of alleles rather than the usual outcome in terms of *observed* counts. In this study, we show analytically and by computer simulation that unequal amplification should be taken into account when testing for differences in allele frequencies between pools, and suggest a simple modification of the standard χ^2 test to control the type I error rate in the presence of experimental error variation. The impact of experimental errors on the power of association studies is shown. *Genet Epidemiol* 24:291–296, 2003. © 2003 Wiley-Liss, Inc.

Key words: DNA pool; association test; SNP; case-control design

Grant sponsor: UK Biotechnology and Biological Sciences Research Council; Grant sponsor: Medical Research Council.
*Correspondence to: Dr. P.M. Visscher, Institute of Cell, Animal, and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK. E-mail: peter.visscher@ed.ac.uk
Received for publication 27 September 2002; Revision accepted 6 December 2002
DOI: 10.1002/gepi.10240

INTRODUCTION

Association studies using DNA pools are a powerful and efficient approach to detect association between a marker allele and disease status, because it reduces the number of genotyping reactions required by a factor of 100–1,000 [Pacek et al., 1993; Shaw et al., 1998; Bader et al., 2001; Sham et al., 2002]. A common observation with the use of DNA pools is that the two alleles at a polymorphic SNP locus are not amplified in equal amounts in heterozygous individuals. In addition, there is experimental error in that there is variation in the estimates of allele frequencies from different pools that are from the same individuals.

The aim of this study was to investigate the impact of additional sources of variation in the estimation of allele frequency on the type I and type II errors in case-control designs, and to propose a new and simple statistical test to analyze association data from DNA pools in the presence of experimental errors.

METHODS

ASSOCIATION STUDY USING OBSERVED ALLELE COUNTS

If N diploid individuals are randomly sampled from a population in Hardy-Weinberg equilibrium, then the sampling variance of the estimate, $\hat{p}$, of the allele frequency is $\mathrm{var}(\hat{p})=p(1-p)/(2N)$. To test the null hypothesis of $p=p_0$, we can use the test statistic $T_{p0}=(\hat{p}-p_0)^2/\mathrm{var}(\hat{p})$. For a large value of N (say, $N>100$), this test statistic is approximately distributed as a χ^2 with 1 degree of freedom under the null hypothesis. In practice, the estimate of the sampling variance is substituted for the true sampling variance, by using $\hat{p}$ instead of p. For a case-control design, the observed allele counts can be summarised in a 2×2 contingency table (see Table I for notation). The standard χ^2 test statistic of independence based upon observed counts (T_{obs1}) can be written as

$$T_{obs1}=(ad-bc)^2(a+b+c+d)/[(a+b)(c+d)(a+c)(b+d)] \quad (1)$$

TABLE I. 2 × 2 Contingency table for SNP-based case-control association study, showing number of observed alleles in each population[a]

	Population		
	Cases	Controls	
SNP allele			
Allele 1	a	b	a+b
Allele 2	c	d	c+d
	a+c	b+d	

[a] $2N_{case}$=a+c; $2N_{control}$=b+d; N_{all1} =a+b; N_{all2}=c+d.

[e.g., Sokal and Rohlf, 1995]. Under the null hypothesis of equal allele frequencies in cases and controls, and for large N_{case} and $N_{control}$ and not too extreme population frequencies, this test statistic is distributed as a χ^2 with 1 degree of freedom. An alternative test statistic is to consider the difference between the allele frequencies from the two groups and the estimated variance of that difference. First, let

$$\hat{p}_{case} = a/(2N_{case}), \hat{p}_{control} = b/(2N_{control}),$$

$$\hat{\mathrm{var}}(\hat{p}_{case}) = \hat{p}_{case}(1-\hat{p}_{case})/(2N_{case}), \text{and}$$

$$\hat{\mathrm{var}}(\hat{p}_{control}) = \hat{p}_{control}(1-\hat{p}_{control})/(2N_{control}).$$

Then, the variance of the difference in estimated allele frequencies is simply,

$$\mathrm{var}(\hat{p}_{case} - \hat{p}_{control}) = \mathrm{var}(\hat{p}_{case}) + \mathrm{var}(\hat{p}_{control}).$$

Analogous to the test statistic for a sample from a single population, we can test the null hypothesis that the frequencies in the two populations are equal by

$$T_{obs2} = (\hat{p}_{case} - \hat{p}_{control})^2 / \hat{\mathrm{var}}(\hat{p}_{case} - \hat{p}_{control}).$$

Asympotically, this test statistic is also distributed as χ^2 with 1 degree of freedom under the null hypothesis of equal allele frequencies. For $N > 100$ and for $0.1 < P < 0.9$, the statistics T_{obs1} and T_{obs2} give virtually identical results, because for these parameters the binomial distribution is well-approximated by a normal distribution. Note that Hardy-Weinberg equilibrium generally does not hold at marker loci that are associated with a disease locus.

ASSOCIATION STUDY USING ESTIMATED ALLELE COUNTS FROM DNA POOLS

There are a number of complications that arise when the allele frequency is estimated from a pool of DNA. Firstly, the estimate of the allele frequency can be biased due to a preferential amplification of one of the alleles, and secondly the estimate of the sample frequency can be imprecise due to unequal amounts of DNA per individual in the pool and due to experimental errors. In this study, we concentrate on the bias and imprecision due to experimental pooling errors, and assume that the errors due to unequal contributions from individuals is negligible. The impact of errors from unequal contributions was examined empirically [Le Hellard et al., 2002] and was found to be negligible.

The output from the PCR analysis is the height of two peaks (A and B) corresponding to two polymorphic alleles at the SNP locus. For heterozygotes, the heights of A and B are not necessarily the same. The frequency in the population of the first allele, corresponding to peak A, is p. Inference about the allele frequency is made from the ratio of the peak heights. Following Hoogendoorn et al. [1999, 2000] and Norton et al. [2002], the ratio of A to B (k=A/B) is estimated for each SNP from a number of independent heterozygotes. For a particular SNP locus, the resulting estimate of k is assumed to be normally distributed, $\hat{k} \sim N(k, \sigma_k^2)$, with $\sigma_k = SE(\hat{k})$. The error in estimating k arises from variation in the quality of the DNA from each heterozygote, and from a pure experimental error attached to each individual analysis. For the purpose of this note, these two sources of error are combined. They could be separated by performing repeated analyses from different heterozygotes. At the population level, p=E(A/B)/(E(A/B) + k). The estimate of the sample allele frequency in the pool is $\tilde{p}$= A/(A + $\hat{k}$B)=(A/B)/(A/B + $\hat{k}$), and the estimated count of alleles is $(2\tilde{p}N)$.

We assume a simple linear model for the sample frequency estimated from the pool,

$$\begin{aligned}\tilde{p} &= \hat{p} + e_k + e_p \\ &= p + e_n + e_k + e_p\end{aligned}$$

with e_n the binomial sampling error, e_k the error due to estimating the correction factor k, and e_p the pool-specific experimental error. The variance of the estimated allele frequency as a function of the variance of $\hat{k}$, $\mathrm{var}(e_k)$, was derived using a first-order Taylor series (Appendix A), and is, approximately,

$$\mathrm{var}(e_k) \approx p^2(1-p)^2\,CV^2(\hat{k})$$

with $CV(\hat{k}) = \sigma_k/k = SE(\hat{k})/k$, i.e., the coefficient of variation of $\hat{k}$. Note that we define CV here as the

standard error relative to the mean, rather than the usual definition as the ratio of the standard deviation and the mean. We further assume that the pool-specific errors (e_p) are normally distributed. Note that the error variation is assumed to be independent of the frequency p, i.e., the experimental noise is assumed to be the same for rare and common alleles. Following these assumptions, the variance of the estimated allele frequency from the pool is

$$\begin{aligned}\mathrm{var}(\tilde{p}) &= \mathrm{var}(\hat{p}) + \mathrm{var}(e_k) + \mathrm{var}(e_p)\\ &\approx p(1-p)/(2N) + p^2(1-p)^2 CV^2(\hat{k})\\ &\quad + \mathrm{var}(e_p).\end{aligned}$$

In summary, we have assumed that there are three potential sources of bias or imprecision: 1) due to sampling a finite number of individuals from a population (the standard sampling error), 2) due to estimating the adjustment factor k, and 3) due to a pool-specific measurement error. Error 1 is reduced by increasing the sample size, error 2 is reduced by using more heterozygotes to estimate k and var(e_k) and/or more replicates from a single heterozygote, and error 3 is reduced by using replicate samples of the pools.

When comparing the frequencies in two pools (e.g., in cases and controls), the variance of the difference in estimated frequency is a function of the difference in population frequency, sample size, the error in estimating k, and the experimental pool error. The source of error due to estimating k will induce a covariance between the estimates from the two pools because the error is the same for both pools. Because the same error in estimating k is made for both pools, the *difference* between the estimates of the frequency in both pools is, to a first-order approximation, negligibly affected (see Appendix A). This was also found empirically by Norton et al. [2002]. The sampling variance of the difference in estimated sample frequencies between the pools is

$$\begin{aligned}\mathrm{var}(\tilde{p}_{case} - \tilde{p}_{control}) &\approx p_{case}(1-p_{case})/(2N_{case})\\ &\quad + p_{control}(1-p_{control})/(2N_{control})\\ &\quad + CV^2(\hat{k})[p_{case}(1-p_{case}) - p_{control}(1-p_{control})]^2\\ &\quad + 2\mathrm{var}(e_p).\end{aligned}$$

Under the null hypothesis that $p_{case}=p_{control}=p$, and equal numbers of individuals in each pool (N), the variance of the difference simplifies to $\mathrm{var}(\tilde{p}_{case} - \tilde{p}_{control}) = p(1-p)/N + 2\mathrm{var}(e_p)$.

One (naive) test statistic is to substitute the estimated allele counts from the pools for the observed counts in Equation (1), and use

$$\begin{aligned}T_{est1} &= (\hat{a}\hat{d} - \hat{b}\hat{c})^2(\hat{a}+\hat{b}+\hat{c}+\hat{d})/\\ &\quad [(\hat{a}+\hat{b})(\hat{c}+\hat{d})(\hat{a}+\hat{c})(\hat{b}+\hat{d})].\end{aligned}$$

The analogous test statistic based on estimated counts and the ratio of the squared differences and a naive estimate of its variance is

$$T_{est2} = (\tilde{p}_{case} - \tilde{p}_{control})^2/\mathrm{var}(\hat{p}_{case} - \hat{p}_{control}).$$

Both of these tests are anticonservative, because the variation due to experimental error is not accounted for properly. Under the null hypothesis of equal allele frequencies in both pools,

$$(\tilde{p}_{case} - \tilde{p}_{control})^2/\mathrm{var}(\tilde{p}_{case} - \tilde{p}_{control}) \sim \chi^2_{(1)}$$

[e.g., Bader et al., 2001; Jawaid et al., 2002]. The expectation of this ratio is 1.0, while the expected value of both test statistics that ignore the extra sources of variation is approximately

$$\begin{aligned}E(T_{est}) &\approx E(\tilde{p}_{case} - \tilde{p}_{control})^2/\mathrm{var}(\hat{p}_{case} - \hat{p}_{control})\\ &= \mathrm{var}(\tilde{p}_{case} - \tilde{p}_{control})/[\mathrm{var}(\hat{p}_{case}) + \mathrm{var}(\hat{p}_{control})]\\ &= [p_{case}(1-p_{case})/(2N_{case})\\ &\quad + p_{control}(1-p_{control})/(2N_{control})\\ &\quad + 2\mathrm{var}(e_p)]/[p_{case}(1-p_{case})/(2N_{case})\\ &\quad + p_{control}(1-p_{control})/(2N_{control})]\\ &= 1 + (\mathrm{var}(e_p)/[2\mathrm{var}(\hat{p}_0)],\end{aligned}$$

with $\hat{p}_0$ the estimate of the allele frequency across the two pools under the null hypothesis, i.e., $\hat{p}_0=(a+b)/(2N_{case}+2N_{control})$, and its variance obtained from the binomial distribution, $\mathrm{var}(\hat{p}_0)=\hat{p}_0(1-\hat{p}_0)/(2N_{case}+2N_{control})$. Under the null hypothesis of equal allele frequencies, the expected value of the test statistic based upon observed counts is $E(T_{obs})=1$. Hence, the test statistic is inflated by the extra source of errors in estimating the allele frequencies. This suggests a simple adjusted test,

$$T^*_{est} = T_{est}\,[2\mathrm{var}(\tilde{p}_0)]/[2\mathrm{var}(\tilde{p}_0) + \mathrm{var}(e_p)],$$

i.e., a shrunk version of the χ^2 test statistic based on estimated counts, with the estimate of the sampling variance of the allele frequency under the null hypothesis obtained from the estimated counts (i.e., $\tilde{p}_0$ replacing $\hat{p}_0$).

SIMULATION

The effect of a larger variance in allele frequencies when dealing with estimated rather than observed counts was investigated using computer simulation, using the above models. A case-control design was simulated by sampling the

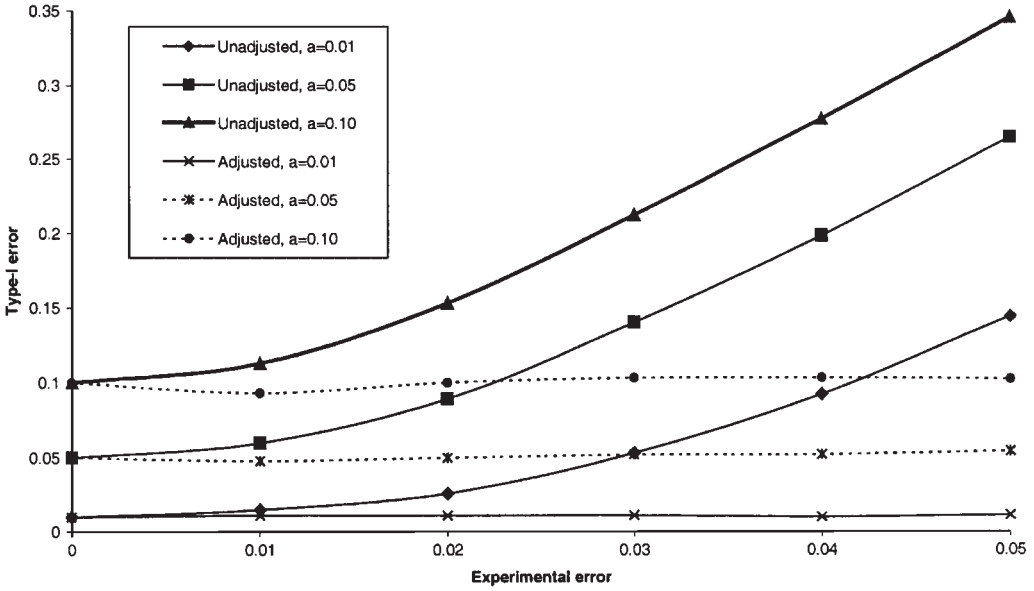

Fig. 1. Empirical type I error rates for unadjusted and adjusted χ^2 tests, for N=100 cases and N=100 controls and P=0.5, from 100,000 replicated Simulations. y-axis, type I error rate; x-axis, experimental error σ_{e_p}.

number of alleles in each group from a binomial distribution. If the adjustment factor k was estimated, the pool frequency before any experimental error was calculated from the sample frequency $\hat{p}$ is

$$p_{pool}=\hat{p}k/[\hat{p}k+\hat{k}(1-p)]$$

with $\hat{k} \sim N(k, var(e_k))$. Finally, the estimate of the pool frequency was calculated as $\tilde{p}=p_{pool}+e_p$, with $e_p \sim N(0, var(e_p))$. Data were simulated either for the null distribution of equal allele frequencies in the pools or for the alternative case when frequencies differed among pools. For each set of parameters, 100,000 simulations were performed.

TABLE II. Power for α=0.05 and 100 cases and 100 controls[a]

σ_e	p(cases)	p(controls)	T_{obs1}	T^*_{est1}
0.01	0.50	0.45	0.17	0.16
		0.40	0.52	0.48
		0.35	0.86	0.83
0.025	0.50	0.45	0.17	0.13
		0.40	0.52	0.38
		0.35	0.86	0.71
0.05	0.50	0.45	0.17	0.09
		0.40	0.52	0.22
		0.35	0.86	0.42

[a]Based on 100,000 replications.

RESULTS

SIMULATION

The impact of estimating k on type I and type II errors was negligible for $CV(\hat{k}) < 0.3$ (results not shown), as predicted, and further results are from simulations in which $\hat{k} = k$. Results from the simulation under the null hypothesis are shown in Fig. 1. Generally, unless the sources of errors are large, the inflation in type I error is small. If the pool-specific error is large $(\text{say}, \sigma_{e_p} > 0.025)$, then the type I error can be inflated substantially. For example, for $\sigma_{e_p} = 0.025$, the type I error is at least doubled relative to the type I error rate on the observed counts. The new test appears to control type I error well, and the behavior of the test statistic is as expected.

Regarding type II error, power is reduced when using the adjusted statistical test relative to the power based on observed counts (Table II). For $\sigma_{e_p} > 0.025$, the reduction in power can be substantial.

DISCUSSION

A new test was proposed to adjust the χ^2 value for the knowledge that counts were estimated and not observed. If the individual pool-specific error

is small ($\sigma_{e_p} < 0.01$), then using either the standard test or the adjusted test makes little difference in inference. However, for large pool-specific errors, the type I error would be inflated substantially if no account was taken of the inflated differences between allele frequencies in the pools. The new test appears to control the type I error well. To achieve an experimental error of 0.01 or less, replicate pools need to be used. If the estimate of between-pool variation in the estimate of the allele frequency is in the range of 0.02–0.05 (SD), then to achieve SE of <0.01, approximately 4–25 replicate pools would give the same power as tests based on observations on individual genotypes. Alternatively, to achieve the same power for direct genotyping as with pooling, the pool sample size must be increased by a factor of $1/[1-2\text{var}(e_p)/\text{var}(\Delta)]$, with var(Δ) the variance of the difference in allele frequencies in the two groups obtained from observed counts. For example, for $\sigma_{e_p}=0.01$ and $\sigma_\Delta = 0.03$ (which corresponds to, for example, 200 cases and 200 controls with frequencies of 0.3 and 0.2, respectively), the sample size of the pool would have to be increased by a factor of $1/(1-0.0002/0.0009)=1.3$, or 30%.

Le Hellard et al. [2002] reported empirical results for pools for five SNPs using three different genotyping technologies. The estimated value of k varied from 0.27–0.95. Using replicate samples of pools with DNA from 96 individuals, the empirical pool-specific experimental error, expressed as the standard deviation of estimates of the sample allele frequency across replicate pools, varied from 0.009–0.135. There was no relationship detected between the size of the pool, in terms of the number of individuals represented in the pool, and the mean or variation in pool-specific errors [Le Hellard et al., 2002]. These results justify the assumptions regarding the range of k-values and σ_{e_p} that were chosen in this study.

An alternative approach to the analysis of pool data would be to fit an overdispersion model in which the pool-specific error is proportional to the binomial sampling error, i.e., $\text{var}(\tilde{p})=cp(1-p)/(2N)$, with c being a constant ($c \geq 1$). The overdispersion parameter c could be estimated from a nested design of population samples and replicated pools within samples.

If the amplification of both alleles is approximately equal in heterozygotes, then a test based on the relative ratios of peaks A and B is equivalent to a test based on observed counts. It might therefore be suggested that adjusting the peak ratio using the factor k is not necessary, and that a statistical test can be performed using the unadjusted peak ratios. However, as shown in Appendix B, even in the absence of any pooling errors this approach should not be used because the behavior of the test statistic depends on the true value of both k and p. In practice, k should be estimated. The simulation results indicated that the precision of estimation does not need to be high. For example, for a $\text{CV}(\hat{k}) < 0.3$, i.e., a scenario where the standard error of the estimate of k is less than 30% of the mean value, the impact on type I error was negligible. However, failing to estimate k, by implicitly assuming that the peak ratio is unity, gives a systematic bias in the test unless the true value is close to unity (Appendix B).

The pool-specific error variance is estimated from replicated pools and needs to be estimated with reasonable accuracy to ensure the correct properties of the proposed test. In practice, this has implications for resource allocation, because a balance needs to be struck between the number of SNPs to be tested and the number of replicate pools per SNP. We used 10 replicates per pool previously, and this appeared to be adequate [Le Hellard et al. 2002].

ACKNOWLEDGMENTS

We thank Andrew Carothers, Ian White, Naomi Wray, Albert Tenesa, and Bill Hill for helpful comments and discussions, and are grateful to our three referees for feedback.

REFERENCES

Bader JS, Bansal A, Sham PC. 2001. Efficient SNP-based tests of association for quantitative phenotypes using pooled DNA. Genescreen 1:143–50.

Hoogendoorn B, Owen MJ, Oefner PJ, Williams N, Austin J, O'Donovan MC. 1999. Genotyping single nucleotide polymorphisms by primer extension and high performance liquid chromatography. Hum Genet 104:89–93.

Hoogendoorn B, Norton N, Kirov G, Williams N, Hamshere ML, Spurlock G, Austin J, Stephens MK, Buckland PR, Owen MJ, O'Donovan MC. 2000. Cheap, accurate and rapid allele frequency estimation of single nucleotide polymorphisms by primer extension and DHPLC in DNA pools. Hum Genet 107:488–93.

Jawaid A, Bader JS, Purcell S, Cherny SS, Sham P. 2002. Optimal selection strategies for QTL mapping using pooled DNA samples. Eur J Hum Genet 10:125–32.

Le Hellard S, Ballereau SJ, Visscher PM, Torrance HS, Pinson J, Morris SW, Thomson ML, Semple CAM, Muir WJ, Blackwood DHR, Porteous DJ, Evans KL. 2002. SNP genotyping on pooled

DNAs: comparison of genotyping technologies and a semi automated method for data storage. Nucleic Acids Res 30:74. 1–10.
Norton N, Williams NM, Williams HJ, Spurlock G, Kirov G, Morris DW, Hoogendoorn B, Owen MJ, O'Donovan MC. 2002. Universal, robust, highly quantitative SNP allele frequency measurement in DNA pools. Hum Genet 110:471–8.
Pacek P, Sajantila A, Syvanen AC. 1993. Determination of allele frequencies at loci with length polymorphism by quantitative analysis of DNA amplified from pooled samples. PCR Methods Appl 2:313–7.
Sham P, Bader JS, Craig I, O'Donovan M, Owen M. 2002. DNA pooling: a tool for large-scale associations studies. Nat Rev Genet 3:862–71.
Shaw SH, Carrasquillo MM, Kashuk C, Puffenberger EG, Chakravarti A. 1998. Allele frequency distributions in pooled DNA samples: applications to mapping complex disease genes. Genome Res 8:111–23.
Sokal RR, Rohlf FJ .1995. Biometrics. New York: W.H. Freeman and Co.

APPENDIX A

POOL SAMPLE FREQUENCY AS A FUNCTION OF ESTIMATING ADJUSTMENT FACTOR K

A first-order Taylor expansion around $\hat{k} = k$ of the pool sample frequency p estimated from k (denoted $\hat{p}_k$) is

$$\hat{p}_k|(\hat{k} = \mathrm{k}, \mathrm{A/B} = \mathrm{E(A/B)}) \approx (\mathrm{A/B})/(\mathrm{A/B} + \mathrm{k}) - (\mathrm{A/B})/(\mathrm{A/B} + \mathrm{k})^2\,(\hat{k} - \mathrm{k})$$

The mean and variance of p as a function of $\hat{k}$ are

$$\mathrm{E}(\hat{p}_k) \approx (\mathrm{A/B})/(\mathrm{A/B} + \mathrm{k}) = \mathrm{p}, \text{and}$$

$$\mathrm{var}(\hat{p}_k) \approx \mathrm{p}^2(1 - \mathrm{p})^2\,\mathrm{var}(\hat{k})/\mathrm{k}^2 = \mathrm{p}^2(1-\mathrm{p})^2\,\mathrm{CV}^2(\hat{k}).$$

A second-order approximation of the mean is

$$\mathrm{E}(\hat{p}_k) \approx \mathrm{p}[1 + (1 - \mathrm{p})^2\,\mathrm{CV}^2(\hat{k})].$$

This expression gives very similar answers to the first-order approximation, unless p is very small (<0.1) and the CV is large (>0.5).

Similarly, the covariance between frequencies in two pools which are estimated with the same estimate of k is, to a first-order approximation

$$\mathrm{cov}(\hat{p}_{case(k)}, \hat{p}_{control(k)}) \approx \mathrm{p_{case}}(1 - \mathrm{p_{case}})\mathrm{p_{control}}\,(1 - \mathrm{p_{control}})\mathrm{CV}^2(\hat{k}).$$

The variance of the difference between $\hat{p}_{case(k)}$ and $\hat{p}_{control(k)}$ is, approximately,

$$\mathrm{var}(\hat{p}_{case(k)} - \hat{p}_{control(k)}) = \mathrm{var}(\hat{p}_{case(k)}) + \mathrm{var}(\hat{p}_{control(k)}) - 2\mathrm{cov}(\hat{p}_{case(k)}, \hat{p}_{control(k)}) \approx [\mathrm{p_{case}}(1 - \mathrm{p_{case}}) - \mathrm{p_{control}}(1 - \mathrm{p_{control}})]^2\,\mathrm{CV}^2(\hat{k}).$$

APPENDIX B

STATISTICAL TEST BASED ON UNADJUSTED RATIOS OF PEAK HEIGHTS

Let R=E(A/B), the ratio of peak heights in a large sample from the population, and p and k the population allele frequency and ratio of peak heights in heterozygotes, respectively. Then R=kp/[1 + (k−1)p], and its estimate is $\hat{R} = k\hat{p}/[1 + (k - 1)\hat{p}]$. Using a first-order Taylor series gives

$$\mathrm{E}(\hat{R})=\mathrm{R} \text{ and } \mathrm{var}(\hat{R})=\{\mathrm{k}/[1 + (\mathrm{k}-1)\mathrm{p}]^2\}^2\,\mathrm{var}(\hat{p}).$$

A test statistic to test $\mathrm{R_t}$=R based on the unadjusted peak ratio, for sample size N, is $\mathrm{T_R}=(\hat{R} - R)^2/\mathrm{var}(\hat{R})$, which has expectation

$$\begin{aligned}\mathrm{E}(\mathrm{T_R}) &\approx \mathrm{var}(\hat{R})/\mathrm{E}(\mathrm{var}(\hat{R}))\\ &= \{(\mathrm{k}/[1 + (\mathrm{k} - 1)\mathrm{p}]^2)^2\,\mathrm{p}(1 - \mathrm{p})/\mathrm{N}\}/\{\mathrm{R}(1 - \mathrm{R})/\mathrm{N}\}\\ &= \mathrm{k}/[1 + (\mathrm{k} - 1)\mathrm{p}]^2.\end{aligned}$$

Hence, the expectation of a naive χ^2 test based on the unadjusted peak ratio depends both on the population allele frequency and on the peak ratio for heterozygotes. Similarly, for testing the difference between peak ratios observed in two pools, the expectation of a test statistic $[(\hat{R}_1 - \hat{R}_2)^2]/[(\hat{R}_1(1 - \hat{R}_1)/\mathrm{N}_1 + \hat{R}_2\,(1 - \hat{R}_2)/\mathrm{N}_2)]$ is approximately k, and so also depends on an unknown parameter. These tests thus have unpredictable properties, and should not be used.