

ARTICLE

Hardy–Weinberg equilibrium in genetic association studies: an empirical evaluation of reporting, deviations, and power

Georgia Salanti¹, Georgia Amountza², Evangelia E Ntzani² and John PA Ioannidis^{*,2,3}

¹MRC Biostatistics Unit, Cambridge University, Cambridge, UK; ²Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece; ³Department of Medicine, Institute for Clinical Research and Health Policy Studies, Tufts-New England Medical Center, Tufts University School of Medicine, Boston, MA, USA

We evaluated the testing and reporting of Hardy–Weinberg equilibrium (HWE) in recent genetic association studies, detected how frequently HWE was violated and estimated the power for HWE testing in this literature. Genetic association studies published in 2002 in *Nature Genetics*, *American Journal of Human Genetics*, and *American Journal of Medical Genetics* were assessed. Data were analyzed on 239 biallelic associations using 154 distinct genotype distribution data sets where HWE could be tested. Any information on HWE was given only for 150 (62.8%) associations (92 (59.7%) data sets). Reanalysis of the data showed significant deviation from HWE in the disease-free controls of 20 associations (13 data sets), but only four of them (two data sets) were admitted in the published articles. Another four deviations (in two data sets) were observed in the combined sample of cases and controls of studies where both cases and controls were diseased, and none were reported in the papers. In all six tested multiallelic associations (six data sets), there was violation of HWE, but this was not admitted in the published articles. Power calculations showed that most studies conforming to HWE simply were largely underpowered to detect HWE deviation; for example, power to detect an inbreeding of magnitude $F=0.10$ exceeded 80% in only 11 (7%) of the data sets being tested. This empirical evidence suggests that, even in high profile genetics journals, testing and reporting for HWE is often neglected and deviations are rarely admitted in the published reports. Moreover, power is limited for HWE testing in most current genetic association studies. *European Journal of Human Genetics* (2005) 13, 840–848. doi:10.1038/sj.ejhg.5201410

Published online 13 April 2005

Keywords: Hardy–Weinberg equilibrium; genetic association; polymorphism; bias; power

*Correspondence: Dr JPA Ioannidis, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina 45110, Greece. Tel: +30 265 109 7807; Fax: +30 265 109 7867; E-mail: jioannid@cc.uoi.gr

This work was supported in part by a PENED grant from the General Secretariat for Research and Technology, Greece and the European Commission.

Received 28 October 2004; revised 16 February 2005; accepted 22 February 2005

Introduction

Empirical evidence suggests that molecular genetic studies sometimes show considerable deficiencies regarding their conduct, analysis, or reporting.^{1,2} It is not known to what extent these deficiencies may underlie some of the failures to replicate and validate postulated associations.³ The validity of genetic association studies depends considerably on the use of appropriate controls. Theoretically, disease-free control groups from outbred populations should follow the Hardy–Weinberg equilibrium (HWE).^{4–6} The

same applies to the combined group of cases and controls from studies where all subjects have a specific disease, for example, studies evaluating different treatment or other outcomes, whenever the disease risk *per se* is not influenced by the evaluated polymorphism. HWE is not simply a theoretical law; deviations can signal important problems, errors or peculiarities in the analyzed data sets.^{4–6} The key inferences from a genetic-association study may be compromised if HWE is violated.

Accumulating evidence suggests that HWE reporting may be suboptimal in non-genetics journals,⁷ but there are no empirical data on reporting of HWE in genetics journals. Moreover, it would be interesting to generate some evidence on the power of currently conducted genetic association studies to detect deviation from HWE.

The aim of this study was to examine the extent to which HWE is estimated and by what means, whether analyses and reporting of HWE-related issues are accurate, whether studies have adequate power to assess HWE, and how deviations thereof are handled in recent genetic association studies published in genetics journals.

Methods

Selection of the genetic association studies

We identified genetic association studies by thorough hand searching of all issues published in the year 2002 in three genetics journals (*American Journal of Human Genetics*, *Nature Genetics*, *American Journal of Medical Genetics*). These journals include the two original research journals with the highest impact factor in the genetics field plus the journal that publishes more genetic association studies than any other.

We included all genetic association studies based on unrelated individuals that assessed at least one association between a particular polymorphism and any multifactorial disease or disease outcome. We excluded studies without controls, family-based linkage studies, studies on monogenic diseases, and studies related to the major histocompatibility complex (human leukocyte antigens). The main analyses were focused on studies and associations thereof where the genotype distribution was provided, so that we could test for HWE. We analyzed separately biallelic and multiallelic loci. Hand searching was performed independently by two investigators. Disagreements in study selection were discussed and consensus was reached using a third investigator as arbitrator.

Data extraction

From each of the eligible studies we extracted data on first author, journal, and whatever selection criteria were mentioned for the control group. Whenever available, the distribution of genotypes in cases and controls was extracted for each studied gene–disease association.

We recorded whether the authors reported anything on HWE. If so, we also recorded whether they specified the group used to test HWE (controls, cases, combined cases and controls); how this testing was done (statistical test and software used); and what results were reported thereof (qualitative statement regarding compliance with or violation of HWE, provision of test statistic or *P*-value, other information). All data extraction was performed independently by two investigators. Discrepancies were discussed and consensus was reached using a third investigator as arbitrator.

Analyses

Separate analyses were performed at the level of gene–disease associations and distinct data sets. The same data set (same distribution of genotypes) might have been used to test several associations. For example, for a specific polymorphism, the same disease-free control group may have been compared against various groups of cases with different diseases or outcomes. Or, the same set of patients with a particular disease may have been tested on whether they had or not a number of different outcomes.

First, we estimated the percentage of tested associations and distinct data sets that provided any information on HWE, those reporting the group used, and those reporting additional information about the testing procedure. Second, we tested for HWE and estimated in how many instances there was deviation from HWE. We evaluated HWE in all disease-free control groups. We also evaluated separately HWE in the combined sample of cases and controls when both cases and controls had a disease (but differed in some outcome, eg schizophrenia with poor vs good treatment response) and the polymorphism did not modulate the overall disease risk. In the presence of a gene–disease association, HWE may be violated in individuals with the disease, so we did not seek conformity with HWE in diseased cases.

Numerous alternatives for testing HWE have been proposed in the literature and frequentist approaches are most common.^{8,9} More efficient Bayesian methods have recently attracted attention.^{10–13} However, these methods are rather sophisticated and not easily implemented by nonstatisticians. Thus, the χ^2 -test remains the most popular option. We chose to evaluate HWE by an exact test.⁸ Given its simplicity, it is a reasonable alternative to the χ^2 -test and can be performed easily for biallelic loci. It is comparable to the χ^2 -test in terms of performance (power and outbreeding detection), but has the advantage of being able to deal with low genotype frequencies, when the χ^2 asymptotic distribution is inadequate.^{8,9} The test assesses the conditional probability to observe the number of homozygotes in the sample for fixed sample size. Given that only a finite number of combinations (and thus a finite number of probabilities) occur, the ‘achievable significance level’ is selected to be as close as possible to

$\alpha = 0.05$. For the multiallelic loci the exact test proposed by Guo and Thompson was applied.¹⁴ We then evaluated the concordance between our estimates derived from the raw data (using the $P = 0.05$ threshold) and the inferences of the primary authors.

Finally, for biallelic loci we estimated for each appropriate group (healthy controls or combined cases and controls with disease) the available power to determine deviations from HWE given the specific characteristics of each of the eligible associations. The empirical power of the exact test was approximated through 10 000 simulations. For the calculations we assumed a parameterization of the Hardy–Weinberg law using the inbreeding coefficient F .^{8,11} Denoting the genotype frequencies for homozygotes as P_{AA} and P_{aa} , and the allele frequencies p_A and p_a , then under HWE we test for

$$F = P_{AA}/p_A + P_{aa}/p_a - 1$$

being zero. We calculated the power at 5% significance level to detect inbreeding levels of $F = 0.05, 0.10, 0.50$ as well as for detecting outbreeding levels of $F = -0.05, -0.10$, and -0.50 . Positive F values reflect an excess of homozygotes while negative F values reflect an excess of heterozygotes compared with those expected under HWE. The allele frequencies for the power calculations were derived from the observed data and when two control groups had been reported, these were merged. Alternative models may also be considered for describing deviation from HWE, but F offers a simple, objective measure of the extent of departure from HWE. It should not be mistakenly inferred as a face-value measure of inbreeding, since it is possible that in most cases causes other than inbreeding are responsible for HWE.

Finally, deviations from HWE affect the type I error for gene–disease associations tested on the level of alleles (per-allele model).^{15,16} Type I error is inflated when the estimated inbreeding coefficient is positive whereas for negative coefficient the error deflates. We calculated what the actual type I error would be for each observed inbreeding coefficient F in our data sets. Calculations are based on Schaid and Jacobsen.¹⁵ The ratio of the variance of the association odds ratio under HWE vs the variance under HWE deviation is given by $1/(1 + F)$.

Analyses were undertaken in R software (R Foundation for Statistical Computing, Vienna, Austria, 2004) using the *genetics* and *gap* package. All P -values are two-tailed.

Results

Database description

We initially identified 85 eligible published reports (*American Journal of Medical Genetics* $n = 58$, *American Journal of Human Genetics* $n = 22$, *Nature Genetics* $n = 5$), assessing 776 genetic associations in total (*American Journal of Medical Genetics* $n = 344$, *American Journal of Human Genetics*

$n = 156$, *Nature Genetics* $n = 276$). For over two-thirds of these associations, no data were available on the genotype distributions, mostly because of studies where a large number of polymorphisms were reported to have been screened for an association without any further detail being provided. Data extraction that would allow HWE calculations (excluding studies where only one allele was found) was feasible in 61 eligible articles (*American Journal of Medical Genetics* $n = 42$, *American Journal of Human Genetics* $n = 17$, *Nature Genetics* $n = 2$ [Supplementary Information]) evaluating in total 245 genetic associations. We first focus on 239 associations studying a biallelic locus in a case–control design of which 183 were based on disease-free controls. There were 154 distinct genotype distribution data sets to test HWE (137 with disease-free controls); the remaining 85 associations were studied using the same data sets, but with different outcomes. In two articles where two control groups were used, we considered the merged sample. Among the 154 distinct data sets, the median sample size was 176 (range from 16 to 4899) and the median minor allele frequency was 23% (interquartile range 10–38%).

Reporting of HWE in the articles

Of the 776 tested associations, any reporting on HWE was made in only 224 (29%). As shown in Table 1 any information on HWE was provided in 63% of the tested associations where genotype data were provided. The appropriate group to apply HWE testing (the controls when disease-free and the combined cases and controls otherwise) was successfully selected by the authors in 51 and 50% of the associations respectively. Information on P -values and analyses used was given in only about a fourth of the associations and software was very uncommonly mentioned. The χ^2 -test was the only test applied, and no information was given on the use of asymptotics or not for inference. Deviations were reported in only seven tested associations and with one exception the authors did not elaborate any further on them. When based on distinct data sets, any information on HWE was given for 92 of the 154 data sets (60%), and deviation from HWE was claimed in five.

Reanalysis of HWE and concordance with reporting

Of the 239 tested associations, there were 20 associations using healthy controls and four associations using diseased controls, where the pertinent group (healthy controls and all subjects, respectively) deviated significantly from HWE. The overall rate of significant deviations from HWE was thus 10%.

The 20 deviations from HWE pertained to 13 distinct data sets of the 137 assessed. For five associations (four data sets), the corresponding articles did not report on HWE for these associations.^{17–20} For another 11 associations (seven data sets),^{17,19,21–23} it was stated that the controls were in

Table 1 Reporting on Hardy–Weinberg equilibrium (HWE)

	ALL STUDIES		Studies with disease-free controls		Studies with diseased controls	
	Associations: 239 N (%)	Data sets: 154 N (%)	Associations: 183 N (%)	Data sets: 137 N (%)	Associations: 56 N (%)	Data sets: 17 N (%)
Any HWE reporting	150 (63)	92 (60)	121 (66)	85 (62)	29 (52)	7 (41)
Reported group(s) tested						
Cases	69 (29)	54 (35)	68 (37)	53 (39)	1 (2)	1 (6)
Controls	95 (40)	67 (44)	94 (51)	66 (48)	1 (2)	1 (6)
Combined	37 (15)	12 (8)	9 (5)	6 (4)	28 (50)	6 (35)
Reported p-values	68 (28)	30 (19)	44 (24)	29 (21)	24 (43)	1 (6)
χ^2 -test	103 (43)	55 (36)	74 (40)	50 (36)	29 (52)	5 (29)
Reported software	19 (8)	17 (11)	18 (10)	16 (12)	1 (2)	1 (6)
Reported deviation ^a	7 (3)	5 (3)	7 (4)	5 (4)	0 (0)	0 (0)
Reported no deviation ^a	111 (46)	65 (42)	87 (48)	61 (45)	24 (43)	4 (24)

^aOn the pertinent group, that is, controls for studies with disease-free controls and combined cases and controls for studies with diseased controls.

HWE, while this claim was not in line with our calculations. We should mention that in five of the seven data sets that reported HWE, the respective articles had two control groups: one of the two control groups and the merged control groups deviated significantly from HWE, while the HWE hypothesis could not be rejected for the second control group. Violation of the equilibrium was admitted in only four associations (two data sets) that we found to deviate.^{21,24}

Among the 163 disease-free controls (124 distinct data sets) where we found no statistically significant deviation from HWE, in three (three distinct data sets)^{25,26} the articles reported that HWE was violated, and in 76 (54 distinct data sets) the articles correctly mentioned that HWE was not violated. For the rest of the papers either the authors have not reported anything about testing, or the results were not clear in their report.

We found statistically significant deviation from HWE in the combined cases and controls in four of 56 associations (two of 17 distinct data sets) where both cases and controls were diseased.^{17,27} The authors did not mention anything about these deviations. In the other 52 associations (15 distinct data sets), there was no deviation from HWE, but this was reported on only 24 associations (four data sets).

In the minority of articles where *P*-values were reported, we observed that these did not correspond well to those obtained from our re-analyses of the data, regardless of whether we applied the exact test or the χ^2 -test (data not shown). Among all 239 associations, 74 comprised data where application of the χ^2 -test using asymptotic inference would not be appropriate due to low frequencies. Among these, 22 (30%) did state that they applied the χ^2 -test and no further information was given regarding the inferences drawn.

Multiallelic loci

All six associations for multiallelic loci revealed significant deviation from HWE upon reanalysis of the data. In four

associations, the articles had stated that they had checked HWE and two of them reported that no violation has been detected.

Power calculations

Across the 154 distinct data sets, the median power (interquartile range) to detect deviations of $F=0.05$, 0.10, and 0.50 was 9% (6–13%), 23% (13–36%) and 100% (94–100%), respectively. Overall only one of the 154 data sets had at least 80% power to detect a statistically significant deviation from HWE with $F=0.05$ and only 11 data sets (7%) had at least 80% power to detect an $F=0.10$. The large majority of the samples were adequately powered for detecting $F=0.50$ (134 (87%) data sets) (Figure 1).

The power was generally more limited for detecting outbreeding (Figure 2). Whereas inbreeding is always detectable, the lowest outbreeding coefficient that can be detected for a given genotype distribution is bounded by $\max\{-p/(1-p), -(1-p)/p\}$ where p is the allele frequency. An outbreeding of $F=-0.05$ can be detected for allele frequencies in the range of 4.8–95.2% and 24 out of 154 data sets had allele frequencies outside this range. For the remaining 130 studies the median power was 7% (interquartile range, 3–12%). For $F=-0.10$, 117 data sets of allele frequencies between 9 and 91% yielded median power of 23% (interquartile range, 10–47%). One and 11 data sets, respectively, again had at least 80% power for detecting such deviations. For $F=-0.50$ the power distribution was nearly dichotomous, with no power for minor allele frequencies below 33% and very high power close to 100% for data sets with minor allele frequencies higher than this threshold.

Power was much higher in data sets that were eventually found to deviate significantly from HWE than in those where the hypothesis of HWE conformity could not be rejected: the median power was 35 vs 8% for $F=0.05$, 88 vs 21% for $F=0.10$, while there was no difference in the medians for detecting $F=0.50$. The median power was 35%

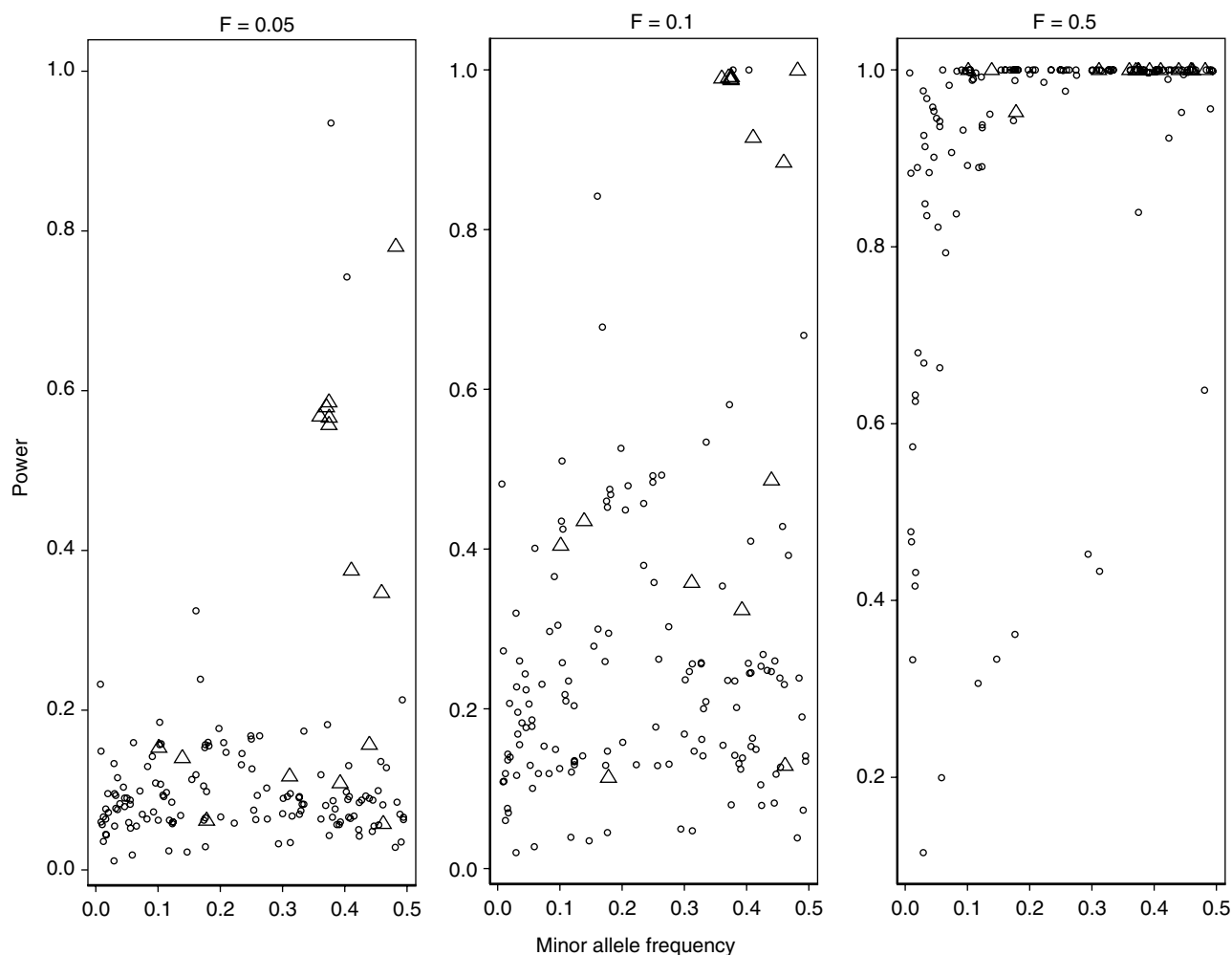


Figure 1 Estimated power of detecting HWE deviations for different positive values of inbreeding coefficient F (including 0.05, 0.10, 0.50). Only samples pertaining to biallelic loci are shown. Both non-diseased controls and diseased cases plus controls samples are shown. Samples where HWE was violated are shown by triangles, while others are shown by circles. The horizontal axis shows the minor allele frequency.

vs 5% for $F = -0.05$ and 87 vs 12% for $F = -0.10$. For $F = -0.50$ the power in the data sets that rejected HWE was maximum (at 100%) vs minimum (nearly 0%) for the other data sets. Power was always higher in the former group (all Mann–Whitney U -test P -values < 0.01). Power also depended on the frequency of the minor allele (Figure 1). Only 26 (65%) of the distinct data sets with minor allele frequency less than 10% had at least 80% power to detect $F = 0.5$ and none had at least 80% power to detect $F = 0.05$ or 0.10. For the same power threshold, none of the data sets with minor allele frequency less than 10% could detect deviance from HWE for any negative F -value.

We also calculated the power of the χ^2 -test for the data sets that fulfilled its requirements (90 data sets). The power was comparable to the power of the exact test, yielding very good agreement (correlation coefficients exceeded 0.99).

Type I error for testing associations on the per-allele model

Among the 15 data sets with statistically significant deviation from HWE (13 data sets of controls and two data sets of cases and controls combined, as described above), the actual type I error for the postulated association was greater than the nominal type I error in five data sets (mean 0.058 for nominal type I error of 0.05) and in 10 data sets it was lower (mean 0.044). In the other data sets, the respective mean type I error was 0.070 for inbreeding and 0.035 for outbreeding. Detailed results are presented in Figure 3.

Discussion

Our empirical evaluation of a large sample of gene–disease associations suggests that reporting of HWE is suboptimal

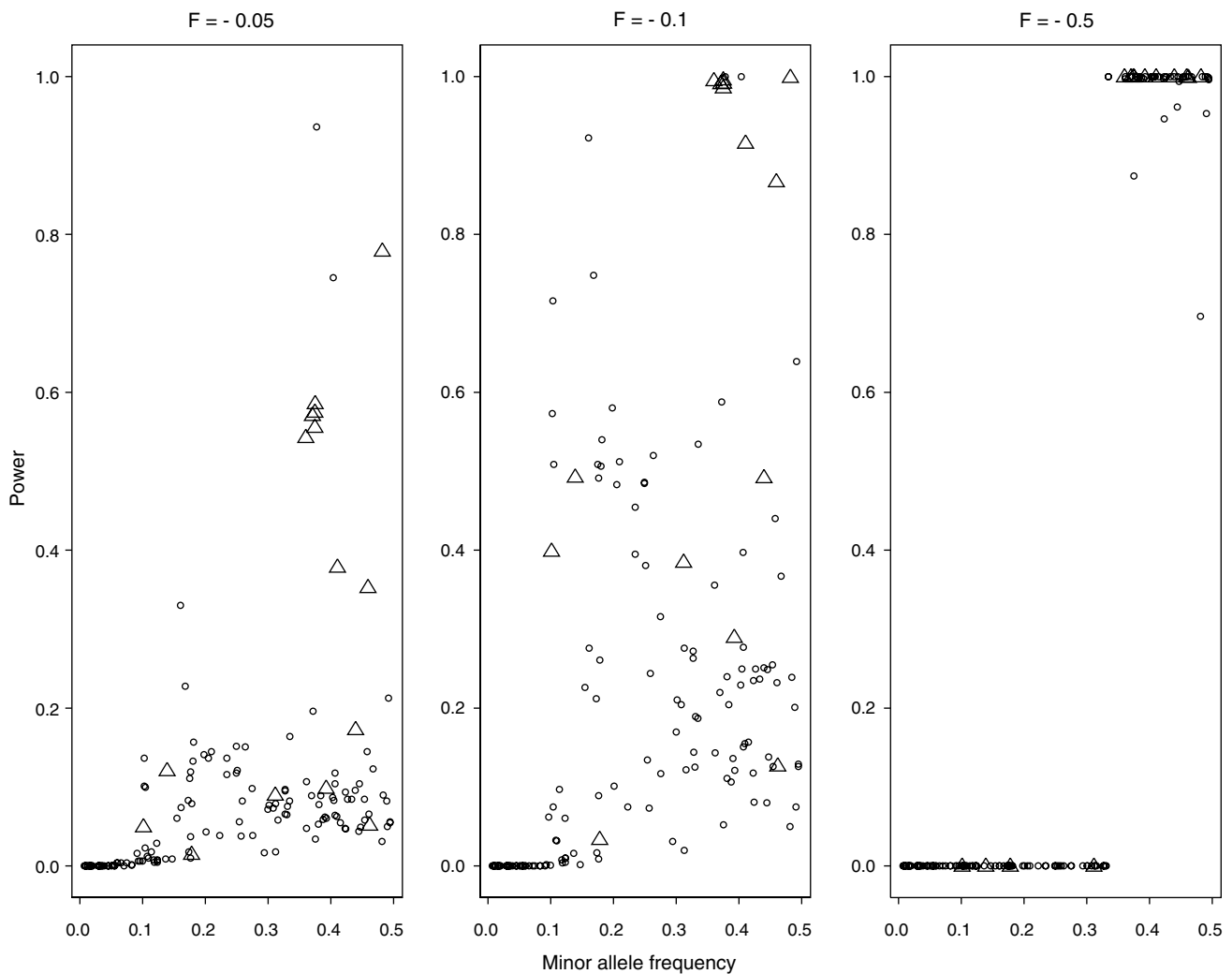


Figure 2 Estimated power of detecting HWE deviations for different negative values of inbreeding coefficient F (including -0.05 , -0.10 , 0.50). Lay out is similar as in Figure 1.

even in high-quality specialized genetics journals. Explicit mention of HWE was made in only 29% of the screened associations. Even when limited to those associations that were scrutinized in more detail and for which genotype distributions were provided, half of the associations were tested without reporting on HWE for the appropriate groups. This does not necessarily mean that the investigators did not check HWE, since they may have done so, but simply failed to report the findings, especially when HWE was not violated. However, it is further disquieting that the reported results of HWE testing were often erroneous. In most of the samples where HWE was actually violated, this was either not mentioned at all in the paper or even a claim was made for HWE conformity. The opposite error (claiming HWE deviation when there was no deviation) was uncommon. Thus, an important reporting

bias seems to exist when handling HWE in genetic association studies.

While the overall rate of deviation from HWE was not high (amounting to about 10% of the tested samples and associations), we should caution that the power to detect deviation was very low in most studies. Power exceeding 80% for detecting $F = 0.10$ or -0.10 was seen in only 7% of the tested associations. Our data suggest that lack of power is a major issue in this literature. It is known that typically the power of the commonly applied HWE tests is limited to address outbreeding in low allele frequencies.⁸ This was also documented in our evaluation, but power was very limited even for detecting inbreeding coefficients of modest magnitude. For minor allele frequencies less than 10%, power was almost never adequate to detect such excesses or deficiencies of homozygotes.

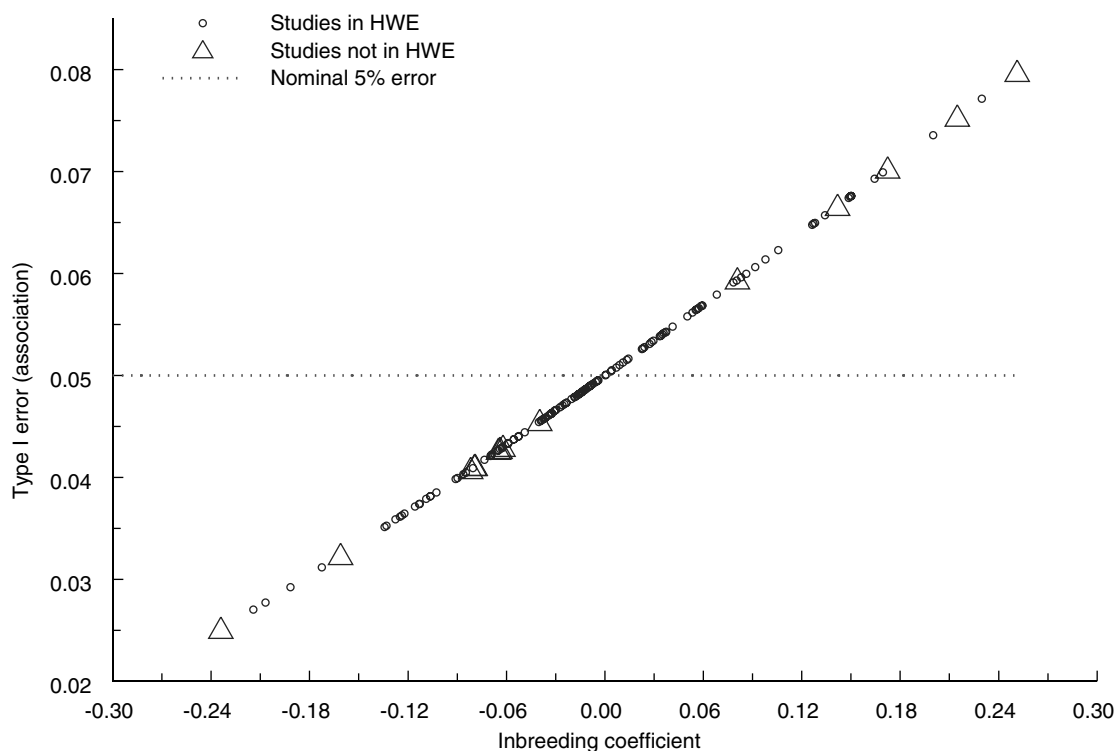


Figure 3 Actual type I error for the per-allele model of genetic association as a function of the inbreeding coefficient F when the nominal type I error under HWE is 0.05. One outlier ($F = -0.73$, type I error < 0.01) is not shown. The data sets shown by triangles are those where there is formally statistically significant violation of HWE.

We should acknowledge that the prime power consideration in genetic association studies should focus on the power to detect an association of plausible magnitude. In this regard, the power to test HWE is probably of rather secondary importance. However, given that most genetic associations have small effect sizes,^{28,29} with odds ratios in the range of 1.1–1.4, modest HWE deviations could considerably affect the inferences of many currently conducted genetic association studies. For example, if there is a recessive model and the control group has an excess or deficit of one group of homozygotes, then this will have a direct impact on the calculation of the odds ratio (the control homozygotes divided by the other genotypes is the denominator of the odds ratio). As we have shown, even for allele-based contrasts (per-allele models), HWE deviations could modestly affect the type I error on some occasions.

In the small number of papers where HWE issue was addressed, some misapplications appeared to occur. The χ^2 -test was often used without justification. Given the small sample sizes and low allele frequencies in many evaluated associations, testing would require an exact test rather than a χ^2 -asymptotic inference. Alternative computational approaches can also be considered.^{10–13,30} Another common misconception is to

test cases for HWE in a study design involving the comparison of diseased cases vs healthy controls for a postulated gene–disease association. In the presence of an association, cases do not need to be in HWE, in fact screening with HWE of data sets of affected individuals has been proposed as a relatively efficient method for detecting gene–disease associations.³¹

Another team of investigators that examined recent publications in diverse medical journals such as *Critical Care Medicine*,³² *Neurology*,³³ *Kidney International*,³⁴ *Gut*,³⁵ *Investigative Dermatology*,³⁶ and *Atherosclerosis*⁷ found that reporting of HWE varies from 20 to 69%. Violations of HWE occurred with a frequency between 10 and 35% and several of these were not admitted by the authors with potentially misleading conclusions for these studies. Another review also found that 12% of data sets did not comply with HWE, but this was acknowledged only by 44% among them.³⁷ The low reporting rate is compatible with our data, although the overall rates of deviation were on the low side of these figures in our empirical evaluation. If not a chance difference, the lower rate of HWE violations in our work may reflect the fact that we targeted recent studies published in specializing genetics journals, where data with HWE deviation may be more likely to capture the attention of editors

and peer-reviewers, as compared with a non-genetics journal. The low rate of acknowledging significant deviations may actually suggest that investigators might have tested for HWE, but felt that acknowledging HWE deviation would create a negative impression about their study. It is also disappointing that only one investigation²⁶ tried to address and discuss why HWE violation might have occurred.

In two-thirds of the originally identified associations in our sample, no genotype data were available so as to allow us testing for HWE. Typically, this included studies that had screened dozens or even more than a hundred polymorphisms for some association, but only those with significant results were reported in any detail without any data on genotypes, let alone HWE, on the others. Although it is difficult to publish detailed genotype data on a very large number of polymorphisms, the availability of electronic databases should allow appropriate recording of this information. Selective reporting may lead to bias in the genetics literature. Some studies targeting only one or a few polymorphisms may also report only on the distribution of alleles or may only report statistics without giving the data from which they are derived. This practice limits the transparency of the data regarding any genotype inferences, including HWE testing.

Conformity with HWE for a locus suggests that several conditions are met including absence of recent mutations and genetic drift and conformance with mendelian segregation and random mating. A nonsignificant HWE test result is simply equivalent to 'non-rejection' of the HWE assumption, but it does not prove that the locus exhibits HWE. HWE is an approximation, because these specific assumptions are rarely perfectly met in human populations plus a large sample is usually required to conform to the 'infinity population' requirement. Deviation from HWE tests may indicate failure in one or more assumptions. For example nonrandom mating may occur with loci related to some special characteristics as deafness and epilepsy. Other explanations as population stratification³⁸ and selection bias are possible. Finally, a probable explanation for deviation from HWE is genotyping error.^{1,37,39,40} HWE deviation may be the strongest and most straightforward hint that genotyping may need to be repeated and double-checked.

Overall, the detection of significant deviation from HWE raises several possibilities for further thinking about a study. Perusing the different options may yield further insights about the data and the population from which they are derived or may lead to more accurate data, if it is found that genotyping error was involved. Departures from HWE may also suggest that allele-based estimates of genetic effects are biased.^{15,16} For all these reasons, HWE testing is a useful analysis that should be routinely and appropriately performed in the setting of genetic association studies.

References

- 1 Bogardus Jr ST, Concato J, Feinstein AR: Clinical epidemiological quality in molecular genetic research: the need for methodological standards. *JAMA* 1999; **281**: 1919–1926.
- 2 Attia J, Thakkinian A, D'Este C: Meta-analyses of molecular association studies: methodologic lessons for genetic epidemiology. *J Clin Epidemiol* 2003; **56**: 297–303.
- 3 Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG: Replication validity of genetic association studies. *Nat Genet* 2001; **29**: 306–309.
- 4 Sham P: *Statistics in human genetics*. London: Arnold Publishers, 2001.
- 5 Khoury MJ, Little J, Burke W: *Human genome epidemiology: a scientific foundation for using genetic information to improve health and prevent disease*. New York: Oxford University Press, 2004.
- 6 Khoury MJ, Beaty TH, Cohen BH: *Fundamentals of genetic epidemiology*. New York: Oxford University Press, 1993.
- 7 Bardoczy Z, Gyorffy B, Kocsis I, Vasarhelyi B: Re-calculated Hardy–Weinberg values in papers published in Atherosclerosis between 1995 and 2003. *Atherosclerosis* 2004; **173**: 141–143.
- 8 Emigh T: A comparison of tests for Hardy–Weinberg equilibrium. *Biometrics* 1980; **36**: 627–642.
- 9 Hernandez JL, Weir BS: A disequilibrium coefficient approach to Hardy–Weinberg testing. *Biometrics* 1989; **45**: 53–70.
- 10 Rogatko A, Slifker MJ, Babb JS: Hardy–Weinberg equilibrium diagnostics. *Theor Popul Biol* 2002; **62**: 251–257.
- 11 Shoemaker J, Painter I, Weir BS: A Bayesian characterization of Hardy–Weinberg disequilibrium. *Genetics* 1998; **149**: 2079–2088.
- 12 Ayres KL, Balding DJ: Measuring departures from Hardy–Weinberg: a Markov chain Monte Carlo method for estimating the inbreeding coefficient. *Heredity* 1998; **80**: 769–777.
- 13 Montoya-Delgado LE, Irony TZ, de B Pereira CA, Whittle MR: An unconditional exact test for the Hardy–Weinberg equilibrium law: sample-space ordering using the Bayes factor. *Genetics* 2001; **158**: 875–883.
- 14 Guo SW, Thompson EA: Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* 1992; **48**: 361–372.
- 15 Schaid DJ, Jacobsen SJ: Biased tests of association: comparisons of allele frequencies when departing from Hardy–Weinberg proportions. *Am J Epidemiol* 1999; **149**: 706–711.
- 16 Sasieni PD: From genotypes to genes: doubling the sample size. *Biometrics* 1997; **53**: 1253–1261.
- 17 Ozaki K, Ohnishi Y, Iida A *et al*: Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat Genet* 2002; **32**: 650–654.
- 18 Shifman S, Bronstein M, Sternfeld M *et al*: A highly significant association between a COMT haplotype and schizophrenia. *Am J Hum Genet* 2002; **71**: 1296–1302.
- 19 Mateo I, Sanchez-Guerra M, Combarros O *et al*: Lack of association between cathepsin D genetic polymorphism and Alzheimer disease in a Spanish sample. *Am J Med Genet* 2002; **114**: 31–33.
- 20 Croen LA, Shaw GM, Barber TC, Baker MM, Finnell RH, Lammer EJ: Apolipoprotein B and apolipoprotein E genotypes and sporadic holoprosencephaly. *Am J Med Genet* 2002; **108**: 75–77.
- 21 Cusin C, Serretti A, Lattuada E, Lilli R, Lorenzi C, Smeraldi E: Association study of MAO-A, COMT, 5-HT2A, DRD2, and DRD4 polymorphisms with illness time course in mood disorders. *Am J Med Genet* 2002; **114**: 380–390.
- 22 Brody LC, Conley M, Cox C *et al*: A polymorphism, R653Q, in the trifunctional enzyme methylenetetrahydrofolate dehydrogenase/methylenetetrahydrofolate cyclohydrolase/formyltetrahydrofolate synthetase is a maternal genetic risk factor for neural tube defects: report of the Birth Defects Research Group. *Am J Hum Genet* 2002; **71**: 1207–1215.
- 23 Villafuerte SM, Del-Favero J, Adolfsson R *et al*: Gene-based SNP genetic association study of the corticotropin-releasing hormone receptor-2 (CRHR2) in major depression. *Am J Med Genet* 2002; **114**: 222–226.

- 24 Heijmans BT, Slagboom PE, Gussekloo J *et al*: Association of APOE epsilon2/epsilon3/epsilon4 and promoter gene variants with dementia but not cardiovascular mortality in old age. *Am J Med Genet* 2002; **107**: 201–208.
- 25 Toyota T, Hattori E, Meerabux J *et al*: Molecular analysis, mutation screening, and association study of adenylate cyclase type 9 gene (ADCY9) in mood disorders. *Am J Med Genet* 2002; **114**: 84–92.
- 26 Tsai MT, Hung CC, Tsai CY *et al*: Mutation analysis of synapsin III gene in schizophrenia. *Am J Med Genet* 2002; **114**: 79–83.
- 27 Poduslo SE, Shook B, Drigalenko E, Yin X: Lack of association of the two polymorphisms in alpha-2 macroglobulin with Alzheimer disease. *Am J Med Genet* 2002; **110**: 30–35.
- 28 Ioannidis JPA: Genetic associations: false or true? *Trends Mol Med* 2003; **9**: 135–138.
- 29 Ioannidis JP, Trikalinos TA, Ntzani EE: 'Racial' differences in genetic effects for complex diseases. *Nat Genet* 2004; **36**: 1312–1318.
- 30 Chakraborty R, Zhong Y: Statistical power of an exact test of Hardy–Weinberg proportions of genotypic data at a multiallelic locus. *Hum Hered* 1994; **44**: 1–9.
- 31 Lee WC: Searching for disease-susceptibility loci by testing for Hardy–Weinberg disequilibrium in a gene bank of affected individuals. *Am J Epidemiol* 2003; **158**: 397–400.
- 32 Nemeth E, Vasarhelyi B, Gyorffy B, Kocsis I: Unreported deviations of genotype distributions from Hardy–Weinberg equilibrium in articles published in *Critical Care Medicine* between 1999 and 2003. *Crit Care Med* 2004; **32**: 1431–1433.
- 33 Kocsis I, Vasarhelyi B, Gyorffy A, Gyorffy B: Reanalysis of genotype distributions published in *Neurology* between 1999 and 2002. *Neurology* 2004; **63**: 357–358.
- 34 Kocsis I, Gyorffy B, Nemeth E, Vasarhelyi B: Examination of Hardy–Weinberg equilibrium in papers of *Kidney International*: an underused tool. *Kidney Int* 2004; **65**: 1956–1958.
- 35 Gyorffy B, Kocsis I, Vasarhelyi B: Biallelic genotype distributions in papers published in *Gut* between 1998 and 2003: altered conclusions after recalculating the Hardy–Weinberg equilibrium. *Gut* 2004; **53**: 614–615.
- 36 Gyorffy B, Kocsis I, Vasarhelyi B: Missed calculations and new conclusions: re-calculation of genotype distribution data published in *Journal of Investigative Dermatology*, 1998–2003. *J Invest Dermatol* 2004; **122**: 644–646.
- 37 Xu J, Turner A, Little J, Bleecker ER, Meyers DA: Positive results in association studies are associated with departure from Hardy–Weinberg equilibrium: hint for genotyping error? *Hum Genet* 2002; **111**: 573–574.
- 38 Cardon LR, Palmer LJ: Population stratification and spurious allelic association. *Lancet* 2003; **361**: 598–604.
- 39 Hosking L, Lumsden S, Lewis K *et al*: Detection of genotyping errors by Hardy–Weinberg equilibrium testing. *Eur J Hum Genet* 2004; **12**: 395–399.
- 40 Gomes I, Collins A, Lonjou C *et al*: Hardy–Weinberg quality control. *Ann Hum Genet*. 1999; **63**: 535–538.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)