

Quantitative Trait Association in Parent Offspring Trios: Extension of Case/Pseudocontrol Method and Comparison of Prospective and Retrospective Approaches

Eleanor Wheeler,¹ and Heather J. Cordell^{2*}

¹The Wellcome Trust Sanger Institute, Cambridge, UK

²Institute of Human Genetics, Newcastle University, UK

The case/pseudocontrol method provides a convenient framework for family-based association analysis of case-parent trios, incorporating several previously proposed methods such as the transmission/disequilibrium test and log-linear modelling of parent-of-origin effects. The method allows genotype and haplotype analysis at an arbitrary number of linked and unlinked multiallelic loci, as well as modelling of more complex effects such as epistasis, parent-of-origin effects, maternal genotype and mother-child interaction effects, and gene-environment interactions. Here we extend the method for analysis of quantitative as opposed to dichotomous (e.g. disease) traits. The resulting method can be thought of as a retrospective approach, modelling genotype given trait value, in contrast to prospective approaches that model trait given genotype. Through simulations and analytical derivations, we examine the power and properties of our proposed approach, and compare it to several previously proposed single-locus methods for quantitative trait association analysis. We investigate the performance of the different methods when extended to allow analysis of haplotype, maternal genotype and parent-of-origin effects. With randomly ascertained families, with or without population stratification, the prospective approach (modeling trait value given genotype) is found to be generally most effective, although the retrospective approach has some advantages with regard to estimation and interpretability of parameter estimates when applied to selected samples. *Genet. Epidemiol.* 31:813–833, 2007. © 2007 Wiley-Liss, Inc.

Key words: family-based; regression; imprinting; TDT

The supplemental materials described in this article can be found at <http://www.interscience.wiley.com/jpages/0741-0395/suppmat>

Contract grant sponsor: Wellcome Trust; Contract grant numbers: 074524 and 068612.

*Correspondence to: Heather Cordell, Institute of Human Genetics, Newcastle University, International Centre for Life, Central Parkway, Newcastle upon Tyne NE1 3BZ, UK. E-mail: heather.cordell@newcastle.ac.uk

Received 24 November 2006; Revised 15 February 2007; Accepted 27 April 2007

Published online 4 June 2007 in Wiley InterScience (www.interscience.wiley.com).

DOI: 10.1002/gepi.20243

INTRODUCTION

Numerous methods have been proposed to test for association between a quantitative trait and a diallelic locus of interest. In a group of unrelated subjects, simple linear regression can be used to relate the quantitative trait phenotype to the genotype. However, this approach can be adversely affected by population stratification [Gauderman, 2003] and hence family-based designs are often preferred. Perhaps the simplest family-based design is to genotype a sample of unrelated phenotyped individuals and their parents, generating a set of parent-offspring trios. Analogous to the transmission/disequilibrium test (TDT) for disease traits [Spielman et al., 1993], tests that are robust to stratification can be derived by focussing on the transmission of the parental alleles to the offspring. If a given marker is not linked to a quantitative trait locus (QTL) (so that the marker alleles are transmitted randomly from parents to offspring), offspring quantitative phenotype is independent of offspring marker

genotype given the parental marker genotypes [Whittaker et al., 2003]. This observation led Whittaker et al. [2003] and Gauderman [2003] to propose a test that is robust to population stratification, by adding terms that code for parental mating type into the linear regression equation. This approach (denoted QTDT_M by Gauderman [2003]) is closely related to tests previously proposed by Allison [1997] and Lunetta et al. [2000] (for details, see Gauderman [2003]).

An alternative approach was proposed by Fulker et al. [1999] and Abecasis et al. [2000]. These authors added terms to the linear regression model to separate out the within and between mating type information. The within mating type test was referred to as the hierarchical QTDT (HQTDT) by Gauderman [2003] and is the method implemented in the QTDT program [Abecasis et al., 2002]. Gauderman [2003] noted that for parent-offspring trios, HQTDT and QTDT_M are virtually identical with regards to inference about the effects of interest (genotype effects on the trait); however HQTDT has the advantage that it has also been extended to apply to general pedigrees [Abecasis

et al., 2002]. Yang et al. [2000] described a similar model to the HQTDT. The differences between the models have little or no effect on the estimates and test of interest [Gauderman, 2003], and the Yang et al. [2000] method was therefore treated as equivalent to HQTDT by Gauderman [2003].

The above models are *prospective* models, that is, they model the quantitative phenotype in terms of the offspring genotype. They all assume that the quantitative traits are normally distributed, or else rely on the central limit theorem. To protect against the effects of possible deviations from either normality or selection on the trait, the HQTDT implemented in the QTDT program can carry out a permutation procedure based on permutation of genotypes to produce an empirical P -value. The connection between maximum likelihood inference under an assumed normal distribution and least-squares regression, however, means that in general we expect these regression-based methods to be reasonably robust to small deviations of the trait from normality, even without use of permutation arguments.

An alternative approach is to use a *retrospective* approach, in which offspring genotype is modelled as a function of quantitative phenotype (possibly given parental genotypes). This approach is more akin to the original TDT of Spielman et al. [1993]. Such an approach provides the rationale for the family-based association test (FBAT) of Laird et al. [2000], although Lange et al. [2002] showed that this retrospective FBAT approach is in fact equivalent to implementing the HQTDT of Abecasis et al. [2000] via a score test rather than via a likelihood ratio test. Kistner and Weinberg [2004, 2005] describe a retrospective approach in which the offspring genotype is modelled as a function of their phenotype and parental genotypes, making no explicit assumptions about the distribution of the quantitative trait. This model, called the quantitative polytomous logistic (QPL) model, can be thought of as an extension of the log-linear model proposed by Weinberg et al. [1998] for qualitative traits.

The log-linear model for qualitative traits proposed by Weinberg et al. [1998] is very similar to the case/pseudocontrol approach for case-parent trios proposed by Cordell and Clayton [2002] (described in more detail by Cordell et al. [2004]). The main difference between the two approaches is that Cordell and Clayton [2002] model offspring genotypes conditional both on parental genotype and ascertainment through the affected offspring, whereas Weinberg et al. [1998] model the frequencies of the 15 possible trio types consisting of offspring genotype and parental mating type. The case/pseudocontrol approach can be thought of as a generalization of the TDT and the approaches of Schaid and Sommer [1993], Schaid [1996] and

Weinberg et al. [1998], generalized to allow the fitting of more complex models where several linked and/or unlinked loci may contribute to disease via a combination of offspring and/or maternal genotype or haplotype effects, parent-of-origin effects and gene-gene or gene-environment interactions. Given the flexibility of the case/pseudocontrol approach, here we extend this approach to deal with quantitative traits, and compare the resulting method to original and extended versions of previously proposed quantitative trait association approaches.

METHODS

THE QTDT_M

Gauderman [2003] and Whittaker et al. [2003] incorporate parental mating type as a fixed effect in a linear regression model. The model is of the form

$$y_i = \alpha_M + \beta_{g_i} + e_i \quad (1)$$

where y_i denotes a quantitative phenotype, g_i the genotype at a particular locus for the i th individual ($g_i = 0, 1, 2$ according to whether the genotype is 1/1, 1/2 (or 2/1) or 2/2), and α_M ($M = 1, \dots, 6$) are mating-type specific intercepts. The residual e_i is assumed to be normally distributed with mean 0 and variance σ^2 . The formulation above suggests that the model is parameterized in terms of six mating-type parameters (the α_M) and three child-genotype parameters (β_0, β_1 and β_2), but, in fact, only two child-genotype parameters are estimable, with one of the genotype categories being chosen as the reference genotype category (the baseline genotype category to which the other genotype effects are compared). For example, if 1/1 is chosen as the reference genotype, β_0 is set equal to 0 and the test of no association between y and g is $H_0: \beta_1 = \beta_2 = 0$. Alternatively, if the heterozygous genotype 1/2 is chosen as the reference genotype, β_1 is set equal to 0 and the test of no association between y and g is $H_0: \beta_0 = \beta_2 = 0$.

Similar to the HQTDT of Abecasis et al. [2000], the QTDT_M draws information from both within and between mating types. The HQTDT models the differences across mating types using a between mating type parameter whereas the QTDT_M uses the multiple fixed intercepts α_M . Gauderman [2003] notes that inferences for the genotype effects are the same using the HQTDT and QTDT_M methods. The differences between the methods are in the estimates and interpretation of the mating-type specific intercepts, which are treated as random effects in the HQTDT and fixed effects in QTDT_M.

THE QPL METHOD

Kistner and Weinberg [2004, 2005] describe a retrospective model in which offspring genotype is

modelled as a function of offspring phenotype and parental genotypes. The model is an extension of the log-linear model proposed by Weinberg et al. [1998] for qualitative traits and is fit using a polytomous logistic model with a generalized logit link function. Assuming parental mating symmetry in the population, there are six distinct parental mating types and the offspring genotype is modelled conditional on the offspring's quantitative trait (y_i) and the parental mating type. Let S_M denote the set of possible offspring genotypes for mating type M . The summation $\sum_{g_i^* \in S_M}$ denotes the restricted summation over the offspring genotypes consistent with S_M . The contribution of a trio to the likelihood is modelled as

$$P(g_i|g_{im}, g_{if}, y_i) = \frac{\exp(\beta''_{g_i} y_i + \alpha''_{Mg_i})}{\sum_{g_i^* \in S_M} \exp(\beta''_{g_i^*} y_i + \alpha''_{Mg_i^*})} \quad (2)$$

where β''_g are parameters representing association between quantitative trait and genotype, and α''_{Mg} are nuisance parameters to account for non-Mendelianism and/or population stratification and depend on both parental mating type and the offspring genotype. In the formulation described by Kistner and Weinberg [2004], these parameters are denoted as β_g and α_{Mg} , but here we instead use the double primed notation (β''_g and α''_{Mg}) to distinguish these parameters from the β_g parameters β_g and α_M used in the prospective formulation of equation (1). Kistner and Weinberg [2004] code the offspring, maternal and paternal genotypes as 0, 1 or 2 depending on the number of 'variant' alleles (here considered to be allele 2) they carry (the results of the test will be the same regardless which allele is considered to be the 'variant'). For the i th trio, let the offspring's quantitative trait value be denoted by y_i . Column 3 of Table I shows the conditional likelihoods for all combinations of parent and offspring genotypes that would result from equation (2) if all the parameters were estimable. Note that there are three child genotype parameters (β''_0, β''_1 and β''_2) and seven nuisance parameters ($\alpha''_{010}, \alpha''_{011}, \alpha''_{110}, \alpha''_{111}, \alpha''_{112}, \alpha''_{121}, \alpha''_{122}$) with α''_{jkl} corresponding to the nuisance parameter for the category with (unordered) parental genotypes j and k and child genotype l . In practice, this model is overparameterized, and Kistner and Weinberg [2004] treat the heterozygous offspring genotype, 1 (1/2 or 2/1), as the reference genotype category. This results in a final model with six estimated parameters ($\beta''_0, \beta''_2, \alpha''_{010}, \alpha''_{110}, \alpha''_{112}$, and α''_{122}), with $\beta''_1, \alpha''_{011}, \alpha''_{111}$ and α''_{121} being set equal to zero, as shown in the fourth column of Table I. An alternative parameterization for the child's genotype parameters (which we shall use in our simulation study) would be to set β''_0 to zero and to estimate β''_1 and β''_2 . Regardless of the parameterization chosen, the categorization into three possible offspring genotype categories means that

there are a maximum of three possible terms in the denominator of equation (2), clearly seen in the columns 3 and 4 of Table I.

THE QCPG METHOD

The method we propose is closely related to the approach by Kistner and Weinberg [2004, 2005], but the parameterization and implementation of the methods are somewhat different. Our approach derives from the case/pseudocontrol method for dichotomous traits [Cordell and Clayton, 2002, Cordell et al., 2004]. This method involves constructing (from a sample of case-parent trios) a sample of cases and matched pseudocontrols. We focus here on the 'conditioning on parental genotypes' (CPG) approach of Cordell et al. [2004], which generates pseudocontrols conditional on the mother's and father's genotypes (and possibly also conditional on some other event ξ , such as phase or parent-of-origin being determinable).

The extension of the CPG method to quantitative traits, here named the quantitative CPG (QCPG), is based on a calculation of the conditional likelihood of the offspring genotypes, conditional on the parental genotypes and the offspring phenotypes. For family i , let g_i, g_{im}, g_{if} be the offspring, maternal and paternal genotypes respectively, and let y_i be the offspring's quantitative trait. Then,

$$\begin{aligned} P(g_i|g_{im}, g_{if}, y_i) &= \frac{P(g_i, g_{im}, g_{if}, y_i)}{P(g_{im}, g_{if}, y_i)} = \frac{P(g_i, g_{im}, g_{if}, y_i)}{\sum_{g_i^*} P(g_i^*, g_{im}, g_{if}, y_i)} \\ &= \frac{P(y_i|g_i, g_{im}, g_{if}) P(g_i|g_{im}, g_{if}) P(g_{im}, g_{if})}{\sum_{g_i^*} P(y_i|g_i^*, g_{im}, g_{if}) P(g_i^*|g_{im}, g_{if}) P(g_{im}, g_{if})} \\ &= \frac{P(y_i|g_i) P(g_i|g_{im}, g_{if})}{\sum_{g_i^* \in S'_M} P(y_i|g_i^*) P(g_i^*|g_{im}, g_{if})} \end{aligned} \quad (3)$$

where $\sum_{g_i^*}$ denotes summation over the four possible offspring genotypes and $\sum_{g_i^* \in S'_M}$ denotes summation over all possible offspring genotypes that could have been transmitted to the offspring given the parental genotypes (the probabilities $P(g_i^*|g_{im}, g_{if}) = 0$ for offspring genotypes that are inconsistent with the parental genotypes). This is of the form of the case/pseudocontrol likelihood for qualitative traits [Cordell and Clayton, 2002] with the offspring's affection status replaced by a quantitative phenotype. The likelihood can be calculated via conditional logistic regression as implemented in standard statistical software. In Appendix A we show that the contribution of a trio to the likelihood may be assumed to be of the form

$$\frac{\exp(\beta'_{g_i} y_i + \alpha'_{Mg_i})}{\sum_{g_i^* \in S'_M} \exp(\beta'_{g_i^*} y_i + \alpha'_{Mg_i^*})}. \quad (4)$$

TABLE I. The conditional likelihoods associated with the QPL model

Parents (g_m, g_f)	Offspring (g)	QPL likelihood	
		Overparameterized model	Six estimable parameters estimated
00	0	1	1
02	1	1	1
22	2	1	1
01	0	$\frac{\exp(\beta''_0 y + \alpha''_{010})}{\exp(\beta''_0 y + \alpha''_{010}) + \exp(\beta''_1 y + \alpha''_{011})}$	$\frac{\exp(\beta''_0 y + \alpha''_{010})}{\exp(\beta''_0 y + \alpha''_{010}) + 1}$
	1	$\frac{\exp(\beta''_1 y + \alpha''_{011})}{\exp(\beta''_0 y + \alpha''_{010}) + \exp(\beta''_1 y + \alpha''_{011})}$	$\frac{1}{\exp(\beta''_0 y + \alpha''_{010}) + 1}$
11	0	$\frac{\exp(\beta''_0 y + \alpha''_{110})}{\exp(\beta''_0 y + \alpha''_{110}) + \exp(\beta''_1 y + \alpha''_{111}) + \exp(\beta''_2 y + \alpha''_{112})}$	$\frac{\exp(\beta''_0 y + \alpha''_{010})}{\exp(\beta''_0 y + \alpha''_{110}) + 1 + \exp(\beta''_2 y + \alpha''_{112})}$
	1	$\frac{\exp(\beta''_1 y + \alpha''_{111})}{\exp(\beta''_0 y + \alpha''_{110}) + \exp(\beta''_1 y + \alpha''_{111}) + \exp(\beta''_2 y + \alpha''_{112})}$	$\frac{1}{\exp(\beta''_0 y + \alpha''_{110}) + 1 + \exp(\beta''_2 y + \alpha''_{112})}$
	2	$\frac{\exp(\beta''_2 y + \alpha''_{112})}{\exp(\beta''_0 y + \alpha''_{110}) + \exp(\beta''_1 y + \alpha''_{111}) + \exp(\beta''_2 y + \alpha''_{112})}$	$\frac{\exp(\beta''_2 y + \alpha''_{112})}{\exp(\beta''_0 y + \alpha''_{110}) + 1 + \exp(\beta''_2 y + \alpha''_{112})}$
12	1	$\frac{\exp(\beta''_1 y + \alpha''_{121})}{\exp(\beta''_1 y + \alpha''_{121}) + \exp(\beta''_2 y + \alpha''_{122})}$	$\frac{1}{1 + \exp(\beta''_2 y + \alpha''_{122})}$
	2	$\frac{\exp(\beta''_2 y + \alpha''_{122})}{\exp(\beta''_1 y + \alpha''_{121}) + \exp(\beta''_2 y + \alpha''_{122})}$	$\frac{\exp(\beta''_2 y + \alpha''_{122})}{1 + \exp(\beta''_2 y + \alpha''_{122})}$

The likelihoods are proportional to $P(g | g_m, g_f, y)$, corresponding to all combinations of (unordered) parent and offspring genotypes.

Here β'_{g_i} represent genotype effects, and α'_{Mg_i} are nuisance parameters modelling non-Mendelianism and population stratification. The likelihood is similar to Kistner and Weinberg's QPL likelihood equation (2), except for the summation in the denominator. Columns 3 and 4 of Table II show the conditional probabilities corresponding to all combinations of parent and offspring genotypes for the QCPCG method. By comparing column 4 of Tables I and II, and ignoring constants of proportionality, it can be seen that the QCPCG likelihood is identical to the QPL likelihood, except for the offspring of two heterozygous parents. With two heterozygous parents, the sum in the denominator of the QCPCG likelihood is over a maximum of four possible offspring genotypes. However, Kistner and Weinberg sum over a maximum of three possible offspring genotypes. For example, for a heterozygous offspring, the contribution to the QPL likelihood is

$$\frac{1}{\exp(\beta''_0 y + \alpha''_{110}) + 1 + \exp(\beta''_2 y + \alpha''_{112})}$$

whereas for the QCPCG method the contribution to the likelihood is

$$\frac{1}{\exp(\beta'_0 y + \alpha'_{110}) + 1 + 1 + \exp(\beta'_2 y + \alpha'_{112})}$$

Essentially, in the QCPCG formulation, we distinguish between the two possible heterozygote offspring genotypes 1/2 and 2/1 in the summation in the denominator (although in practice – assuming no parent-of-origin effects – the likelihood will be identical regardless of whether the observed offspring has genotype 1/2 or 2/1). As a result of the different likelihood formulations, interpretation of the nuisance parameters is different for the QCPCG method and Kistner and Weinberg's QPL. The difference is most noticeable under the null with no population stratification or non-Mendelianism. In Appendix A we show that, for the QCPCG method, the true values of the nuisance parameters α' under the null with no population stratification or non-Mendelianism equal zero, whereas in Appendix B we show that for Kistner and Weinberg's QPL method, the true values of the nuisance parameters α'' are non-zero. Provided all six estimable parameters (two offspring genotype effects and four

TABLE II. The conditional likelihoods associated with the QCPG model

Parents (g_m, g_f)	Offspring (g)	QCPG likelihood	
		Overparameterized model	Six estimable parameters estimated
00	0	1	1
02	1	1	1
22	2	1	1
01	0	$\frac{\exp(\beta'_0 y + \alpha'_{010})}{2 \exp(\beta'_0 y + \alpha'_{010}) + 2 \exp(\beta'_1 y + \alpha'_{011})}$	$\frac{\exp(\beta'_0 y + \alpha'_{010})}{2 \exp(\beta'_0 y + \alpha'_{010}) + 2}$
	1	$\frac{\exp(\beta'_1 y + \alpha'_{011})}{2 \exp(\beta'_0 y + \alpha'_{010}) + 2 \exp(\beta'_1 y + \alpha'_{011})}$	$\frac{1}{2 \exp(\beta'_0 y + \alpha'_{010}) + 2}$
11	0	$\frac{\exp(\beta'_0 y + \alpha'_{110})}{\exp(\beta'_0 y + \alpha'_{110}) + 2 \exp(\beta'_1 y + \alpha'_{111}) + \exp(\beta'_2 y + \alpha'_{112})}$	$\frac{\exp(\beta'_0 y + \alpha'_{110})}{\exp(\beta'_0 y + \alpha'_{110}) + 2 + \exp(\beta'_2 y + \alpha'_{112})}$
	1	$\frac{\exp(\beta'_1 y + \alpha'_{111})}{\exp(\beta'_0 y + \alpha'_{110}) + 2 \exp(\beta'_1 y + \alpha'_{111}) + \exp(\beta'_2 y + \alpha'_{112})}$	$\frac{1}{\exp(\beta'_0 y + \alpha'_{110}) + 2 + \exp(\beta'_2 y + \alpha'_{112})}$
	2	$\frac{\exp(\beta'_2 y + \alpha'_{112})}{\exp(\beta'_0 y + \alpha'_{110}) + 2 \exp(\beta'_1 y + \alpha'_{111}) + \exp(\beta'_2 y + \alpha'_{112})}$	$\frac{\exp(\beta'_2 y + \alpha'_{112})}{\exp(\beta'_0 y + \alpha'_{110}) + 2 + \exp(\beta'_2 y + \alpha'_{112})}$
12	1	$\frac{\exp(\beta'_1 y + \alpha'_{121})}{2 \exp(\beta'_1 y + \alpha'_{121}) + 2 \exp(\beta'_2 y + \alpha'_{122})}$	$\frac{1}{2 + 2 \exp(\beta'_2 y + \alpha'_{122})}$
	2	$\frac{\exp(\beta'_2 y + \alpha'_{122})}{2 \exp(\beta'_1 y + \alpha'_{121}) + 2 \exp(\beta'_2 y + \alpha'_{122})}$	$\frac{\exp(\beta'_2 y + \alpha'_{122})}{2 + 2 \exp(\beta'_2 y + \alpha'_{122})}$

The likelihoods are proportional to $P(g | g_m, g_f, y)$, corresponding to all combinations of (unordered) parent and offspring genotypes.

nuisance parameters) are freely estimated during likelihood maximization, inference for the parameters of interest (β''_0 and β''_2 , or β''_1 and β''_2 , depending on which genotype category is chosen as reference) should not be affected by this result. However, the QPL result is slightly counter-intuitive, as one would generally expect that parameters that are specifically included in the likelihood to model certain effects (such as population stratification or non-Mendelianism) would take the value zero (i.e. be removable from the likelihood) when these effects do not, in fact, exist.

An additional difference between the QCPG and QPL arises with regard to the number of nuisance parameters estimated (see Appendix C).

Liu et al. [2002] described a method that is closely related to the QCPG. Although the nuisance parameters are different, the likelihood is essentially of the same form. Rather than having six possible parental mating type parameters (α_M), Liu et al. [2002] have a single baseline parameter, α , and a number of additional parameters for the clusters of individuals whose trait differs from the population mean due to population stratification, δ_i . However, without knowing the underlying population stratification, the δ_i parameters are unknown and cannot be estimated from the data. Liu et al.

[2002] avoid having to estimate these nuisance parameters by showing that even in the presence of unobservable population stratification, it is still valid to test the null of no genetic effect via a score test, since population stratification has no effect on the null distribution of the test. Gauderman [2003] refers to the method of Liu et al. [2002] as the retrospective QTDT (RQTDT). To implement this method via a likelihood ratio test, Gauderman assumed that the quantitative trait follows a normal distribution with mean $\alpha + \beta_{g_i}$ and variance σ^2 , without consideration of the δ_i parameters. In this implementation, α does not use information on the parental genotypes to model population stratification, although some information on the parental genotypes is still incorporated via the genotypes of the offspring and the pseudocontrols.

EXTENSION OF QCPG TO MULTI-LOCUS HAPLOTYPES

Cordell et al. [2004] showed that the case/pseudocontrol approach can easily be extended to fit models for parent-of-origin effects, multiallelic markers, multiple linked loci in multiple unlinked regions, and gene-gene and gene-environment

interactions, via an adjustment to the conditioning argument that results in differing numbers of pseudocontrols depending on the model being fitted. Here we extend this approach to quantitative traits. Consider models in which the genotype effects depend only on child's phased genotype. Define g_i, g_{im}, g_{if} as the offspring, maternal and paternal phase-known genotypes respectively, and y_i as the offspring's quantitative trait. The likelihood is very similar to that in equation (3) but we define an event ξ as the event that the set of transmitted and untransmitted haplotypes from the parents can be deduced. The contribution to the conditional likelihood is

$$\begin{aligned} P(g_i|g_{im}, g_{if}, y_i, \xi) &= \frac{P(g_i, g_{im}, g_{if}, y_i, \xi)}{P(g_{im}, g_{if}, y_i, \xi)} = \frac{P(g_i, g_{im}, g_{if}, y_i, \xi)}{\sum_{g_i^*} P(g_i^*, g_{im}, g_{if}, y_i, \xi)} \\ &= \frac{P(y_i|g_i, g_{im}, g_{if}, \xi) P(g_i|g_{im}, g_{if}, \xi) P(g_{im}, g_{if}, \xi)}{\sum_{g_i^*} P(y_i|g_i^*, g_{im}, g_{if}, \xi) P(g_i^*|g_{im}, g_{if}, \xi) P(g_{im}, g_{if}, \xi)} \\ &= \frac{P(y_i|g_i) P(g_i|g_{im}, g_{if}, \xi)}{\sum_{g_i^* \in S_M} P(y_i|g_i^*) P(g_i^*|g_{im}, g_{if}, \xi)} \end{aligned}$$

where $\sum_{g_i^*}$ denotes summation over all possible offspring genotypes and $\sum_{g_i^* \in S_M}$ denotes summation over all possible offspring genotypes that could have been transmitted to the offspring given the parental genotypes. Under Mendelian inheritance the probabilities $P(g_i^*|g_{im}, g_{if}, \xi)$ are equal for all $g_i^* \in \{S_M \cap G_\xi\}$ and equal zero otherwise, where G_ξ denotes the set of offspring genotypes determined by ξ . Then the contribution to the likelihood under Mendelian inheritance is given by

$$P(g_i|g_{im}, g_{if}, y_i, \xi) = \frac{P(y_i|g_i)}{\sum_{g_i^* \in \{S_M \cap G_\xi\}} P(y_i|g_i^*)}$$

Note that, provided the models that are to be fitted do not depend on phase, one could also use the QCPCG for analysis of unphased multilocus genotype data, in the same way that the CPG method can be used for unphased genotype data [Cordell et al., 2004].

EXTENSIONS FOR MATERNAL GENOTYPE AND PARENT-OF-ORIGIN EFFECTS

Kistner et al. [2006] proposed an extension to the QPL approach to allow testing for maternally mediated effects and parent-of-origin effects. The likelihood factors into two parts. The first factor tests for genotype effects in the offspring and can be modelled using the original QPL method. The second factor tests for maternal genotype or parent-of-origin effects via a logistic regression model.

Maternal genotype effects are incorporated by modelling the probability that the mother has more copies of the variant allele than the father for each mating type. Parent-of-origin effects are incorporated by additionally including a binary indicator variable, indicating whether the offspring inherited only one copy of the variant allele. This implies the child is heterozygous, and since the mother has more copies of the variant allele than the father, the variant allele must have been inherited from the mother.

We may also extend the QCPCG method to allow for maternal genotype and parent-of-origin effects. The 'conditioning on exchangeable parental genotypes' (CEPG) method [Cordell et al., 2004] is an extension of the CPG approach to detecting parent-of-origin or maternal genotype effects by assuming exchangeability of parental genotypes. The method conditions on the set of parental genotypes but not on their order, generating additional pseudocontrols constructed by exchanging the genotypes of the mother and father. Here we extend this approach to quantitative traits. Maternal genotype effects are defined to be the direct effect of the maternal genotype on the offspring's quantitative trait, and parent-of-origin effects are defined (as in Weinberg et al. [1998]) to allow the offspring's quantitative trait to vary according to the parental origin of the variant allele, if present. Like Cordell et al. [2004], we introduce an additional conditioning event ξ corresponding to the event that parent-of-origin and maternal genotype can be deduced in the trio. For quantitative trait CEPG method (denoted here QCEPG), the contribution of a trio to the QCEPG likelihood is

$$\begin{aligned} P(g_i, g_{im}, g_{if}|\{g_{im}, g_{if}\}, y_i, \xi) &= \frac{P(g_i, g_{im}, g_{if}, \{g_{im}, g_{if}\}, y_i, \xi)}{P(\{g_{im}, g_{if}\}, y_i, \xi)} \\ &= \frac{P(g_i, g_{im}, g_{if}, \{g_{im}, g_{if}\}, y_i, \xi)}{\sum_{g_i^*, g_{im}^*, g_{if}^*} P(g_i^*, g_{im}^*, g_{if}^*, \{g_{im}, g_{if}\}, y_i, \xi)} \\ &= \frac{P(y_i|g_i, g_{im}, g_{if}, \xi) P(g_i, g_{im}, g_{if}|\{g_{im}, g_{if}\}, \xi) P(\{g_{im}, g_{if}\}, \xi)}{\sum_{g_i^*, g_{im}^*, g_{if}^*} P(y_i|g_i^*, g_{im}^*, g_{if}^*, \xi) P(g_i^*, g_{im}^*, g_{if}^*|\{g_{im}, g_{if}\}, \xi) P(\{g_{im}, g_{if}\}, \xi)} \\ &= \frac{P(y_i|g_i, g_{im}, g_{if}, \xi)}{\sum_{g_i^*, g_{im}^*, g_{if}^* \in G_\xi} P(y_i|g_i^*, g_{im}^*, g_{if}^*, \xi)} \end{aligned} \tag{5}$$

where $\{g_{im}, g_{if}\}$ denotes the unordered set of parental genotypes and the final restricted sum is over the possible trios in which parent-of-origin (as well as maternal genotype) are deducible. Unlike Kistner and Weinberg's extension of the QPL method, the QCEPG method does not involve factoring the likelihood. In addition, the null hypothesis of no parent-of-origin effects considers the transmission of variant alleles from the mother to all offspring, not only those who are heterozygous. The contribution

of the likelihood (see Appendix D) is assumed to be of the form

$$\frac{\exp((\beta'_{g_i} + \beta'_{g_{im}} + \beta_{I_m})y_i + \alpha'_{g_{im}g_{if}g_i})}{\sum_{g_i^* g_{im}^* g_{if}^* \in G_{\xi}} \exp((\beta'_{g_i^*} + \beta'_{g_{im}^*} + \beta_{I_m^*})y_i + \alpha'_{g_{im}^* g_{if}^* g_i^*})}. \quad (6)$$

The offspring genotype effects are denoted by β'_{g_i} , the maternal genotype effects by $\beta'_{g_{im}}$ and parent-of-origin effects by β_{I_m} , where I_m is an indicator of whether an offspring inherits a variant allele from the mother. The nuisance parameters $\alpha'_{g_{im}g_{if}g_i}$ depend on the offspring genotype and the genotypes of the parents. For the QCPG, in the absence of maternal genotype or parent-of-origin effects, the nuisance parameters only depended on the parental mating type and the offspring genotype. However, in the QCEPG, the maternal genotype is of interest and needs to be included. No additional information is gained by incorporating the parent-of-origin indicator into the nuisance parameter since once the phase-known genotypes of the parents and offspring are specified, and conditional on the fact that parent-of-origin can be deduced, parent-of-origin is also established. If only maternal genotype effects are of interest then the parent-of-origin indicator can simply be removed from the model and the nuisance parameters remain the same. However, if only parent-of-origin effects are of interest then the nuisance parameters are of the form $\alpha'_{Mg_i^* I_m}$, which depend on the parental mating type, the offspring genotype and the parent-of-origin indicator.

The QTDT_M method can also be extended to include maternal genotype effects, $\beta_{g_{im}}$, and parent-of-origin effects, β_{I_m} , through fitting the linear regression model

$$y_i = \alpha_M + \beta_{g_i} + \beta_{g_{im}} + \beta_{I_m}.$$

Trios in which parent-of-origin can be resolved can be found by first generating a case/pseudocontrol dataset as described in Cordell et al. [2004], specifying that the parental genotypes are exchangeable and parent-of-origin can be resolved. For the QTDT_M method, only the original offspring (the 'case') from the case/pseudocontrol dataset (together with information about maternal genotype and parent-of-origin status) is used in the prospective likelihood, whereas in the QCEPG method, the full set of cases and pseudocontrols is required for the retrospective likelihood.

SIMULATION STUDY

SINGLE-LOCUS SIMULATIONS

Simulations were performed to investigate the power and properties of the various methods described. Initially, a single diallelic QTL locus was

considered. One thousand replicates of data were generated, each consisting of a number of genotyped trios (i.e. a single offspring with a quantitative trait and both parents). Bias in the resulting parameter estimates, 95% confidence intervals, power and type I error were examined. A method that performs well would be expected to give unbiased parameter estimation and to show approximately 95% confidence interval coverage. The importance of the nuisance parameters in the retrospective models was also investigated by examining the estimates obtained when they are removed from the model and also when the offspring genotype is used as a substitute.

For the single-locus model, six generating scenarios were considered as shown in Table 1 (**online**). Three different sampling schemes were employed: random sampling, one-tail sampling from the upper tail of the offspring trait distribution and two-tailed sampling from the upper and lower tails of the offspring trait distribution. Under random sampling, 500 parent-offspring trios were simulated per replicate where the offspring's quantitative trait was drawn from a normal distribution with genotype mean and standard deviation as shown in Table 1 (**online**). Population stratification was simulated by combining data in different proportions from two subpopulations, each of which was in Hardy-Weinberg equilibrium and showed random mating. The subpopulations had different allele frequencies and mean quantitative trait values, producing a spurious correlation between the quantitative trait and genotypes when the populations are combined. Under selected sampling from the extremes, 5,000 trios were generated per replicate, from which a subset were selected for analysis. For the two-tailed sampling scheme, 500 trios were selected from the 5,000 (i.e. the top and bottom 5% of the trait distribution). For the one-tail sampling scheme, we selected 1,000 trios from the 5,000 (i.e. the top 20%), as convergence problems were encountered when using only 500 trios under this sampling scheme.

Table III shows results for the first three scenarios with no population stratification and where the trios were randomly sampled. Under the null, all the methods gave unbiased estimates and reasonable confidence intervals, except for the QPL method where the nuisance parameters have been removed. This is expected since under the null, the α'' parameters in the QPL are nonzero and so their removal affects the resulting β'' estimates. Similarly, under the first alternative model (Alt 1) all the methods performed well except the QPL and QCPG methods where the nuisance parameters have been removed. The retrospective models in which the nuisance parameters have been replaced by the offspring genotype parameters give β' and β''

TABLE III. True and estimated means, standard deviations (SD) and coverage (CI) of the 95% confidence intervals for the single locus simulations with random selection and no population stratification

Method	Parameter	Models												
		Null				Alt 1				Alt 2				
		True mean	Mean	(SD)	CI	True mean	Mean	(SD)	CI	True mean	Mean	(SD)	CI	
Linear regression	Constant	-	1.99	(0.46)	-	-0.01	(0.46)	-	-0.01	(0.46)	-	-0.01	(0.46)	-
	β_1	0.00	0.05	(0.50)	0.96	1.05	(0.50)	0.96	1.05	(0.50)	0.96	1.05	(0.50)	0.96
	β_2	0.00	0.04	(0.48)	0.96	2.04	(0.48)	0.96	2.04	(0.48)	0.96	2.04	(0.48)	0.96
QTDT _M	β_1	0.00	0.08	(0.60)	0.95	1.08	(0.60)	0.95	1.08	(0.60)	0.95	1.08	(0.60)	0.95
	β_2	0.00	0.07	(0.63)	0.95	2.07	(0.63)	0.95	2.07	(0.63)	0.95	2.07	(0.63)	0.95
QCPG	β'_1	0.00	0.02	(0.16)	0.95	0.27	(0.17)	0.96	0.27	(0.17)	0.96	0.27	(0.17)	0.95
	β'_2	0.00	0.02	(0.16)	0.95	0.51	(0.17)	0.97	0.51	(0.17)	0.97	0.51	(0.17)	0.96
QPL	β'_1	0.00	0.02	(0.16)	0.95	0.27	(0.17)	0.96	0.27	(0.17)	0.96	0.27	(0.17)	0.95
	β'_2	0.00	0.02	(0.16)	0.95	0.51	(0.17)	0.97	0.51	(0.17)	0.97	0.51	(0.17)	0.96
QCPG α 's removed	β'_1	0.00	0.02	(0.11)	0.96	0.26	(0.15)	0.96	0.26	(0.15)	0.96	0.26	(0.15)	0.96
	β'_2	0.00	0.01	(0.11)	0.95	0.40	(0.16)	0.88	0.40	(0.16)	0.88	0.40	(0.16)	0.96
QPL α 's removed	β'_1	0.00	0.10	(0.10)	0.87	0.27	(0.14)	0.97	0.27	(0.14)	0.97	0.27	(0.14)	0.97
	β'_2	0.00	0.07	(0.10)	0.92	0.40	(0.14)	0.88	0.40	(0.14)	0.88	0.40	(0.14)	0.97
QCPG α 's replaced by offspring genotype (ξ)	β'_1	0.00	0.02	(0.15)	0.96	0.27	(0.16)	0.97	0.27	(0.16)	0.97	0.27	(0.16)	0.96
	β'_2	0.00	0.01	(0.16)	0.95	0.50	(0.16)	0.96	0.50	(0.16)	0.96	0.50	(0.16)	0.96
QPL α 's replaced by offspring genotype (ξ)	β'_1	0.00	0.02	(0.15)	0.96	0.24	(0.16)	0.97	0.24	(0.16)	0.97	0.24	(0.16)	0.96
	β'_2	0.00	0.01	(0.15)	0.96	0.47	(0.16)	0.96	0.47	(0.16)	0.96	0.47	(0.16)	0.96

The simulation parameters are as shown in Table 1 (online supplementary materials).

estimates very close to the true means and reasonable coverage.

The results for the three scenarios in the presence of population stratification and under random sampling are shown in Table IV. Simple linear regression showed the expected bias in the estimates of β and poor coverage of the estimated 95% confidence intervals since the population stratification is not accounted for in the method. Under the first null model, as in the case without population stratification (Table III), the QPL method with the α' parameters removed did not perform well (coverages 0.87 and 0.92 instead of 0.95). The remainder of the methods perform well under both null models, even under population stratification. Substitution of the four nuisance parameters by the offspring genotype in the retrospective methods appears to account sufficiently for the population stratification. Under the alternative with population stratification, only the prospective QTDT_M method produced unbiased estimates and correct coverage. The retrospective methods (QCPG and QPL) produced biased estimates, as expected (see Appendices A and C).

Parameter estimates under one-tail selected sampling are shown in online Tables 2 and 3 (**online**). The results under the null (both with and without population stratification) are the same as those found in the unselected case. Under the alternative with no population stratification, both prospective models (simple linear regression and QTDT_M) show biased estimates and incorrect coverage of the 95% confidence intervals. This is because the methods cannot account for the selection on quantitative trait value. By conditioning on the trait values, the retrospective models should be robust to selection. However, Table 2 (**online**) suggests that these methods are producing biased estimates. By looking at the median genotype effect estimates (data not shown), we found that the bias is due to a small number of outlying observations. The medians for the QCPG and QPL methods (with the true nuisance parameters and with the nuisance parameters replaced by the offspring genotype) are very close to the true means. Under the alternative with population stratification, all methods performed poorly, producing biased estimates and incorrect coverage. Here, the prospective model QTDT_M fails since it cannot account for selection on quantitative trait value and the retrospective models, QCPG and QPL, fail to estimate the nuisance parameters under the alternative with population stratification. Similar results were observed using the two-tailed sampling scheme (Tables 4 and 5 (**online**)). Without population stratification (Table 5 (**online**)), it can be seen that the bias in the estimates using simple linear regression is not as great under two-tailed sampling

as found when sampling only from the upper tail of the trait distribution (note that under the alternative, the assumption of homoscedasticity of the residuals is violated under the one-tailed sampling scheme).

Powers/type I errors are shown in Table 6 (**online**). Since the powers to achieve P value of 0.001 for the different methods are all 1.0 under two-tailed sampling, we also investigated the power to achieve a more stringent significance level in this case. Under the null with no population stratification, removal of the nonzero nuisance parameters in the QPL method generates a bias in the estimates and hence increased type I error rates, most clearly seen in Table 6 (**online**) for the random and one-tail sampling schemes. The remaining methods all have type I error rates close to or less than the critical values. Highest power to detect a genotypic effect is seen with the linear regression method for the random and two-tailed sampling schemes and with the QCPG method with the α' parameters removed for the one-tail sampling scheme (powers are meaningless for the QPL method with no α' parameters since the type I errors are incorrect). In all cases, the highest powers to detect a genotypic effect are seen for the two-tailed selected sampling scheme, selecting from the upper and lower tails of the offspring trait distribution. In contrast, selection from only the upper tail of the offspring trait distribution actually decreases the power to detect a genetic effect compared to the random sampling scheme, despite having the largest sample size, except for the QCPG method with the α' parameters removed. These results also show that, although under the alternative with selected sampling the QTDT_M method showed biased estimates and poor coverage of the 95% confidence intervals, the method can still be used to test for a genetic effect, as the type 1 error is correct. In fact, the large bias in the estimates seen when using a two-tailed sampling scheme actually increases the power to detect an effect compared to random sampling, although this power increase may also be due to the fact that the selected subjects carry more information, since they are concentrated at the extremes of the trait distribution.

Under population stratification, Table 6 (**online**) shows that for all sampling schemes the simple linear regression method has increased type I error rates. Since linear regression cannot account for the population substructure, the resulting bias in the estimates generates a large number of false-positive associations. The QPL method with the α' parameters removed also has type I errors larger than the nominal values, particularly when selecting from the upper tail of the offspring trait distribution, as found in the case of no population stratification. The type I errors for the QTDT_M method under population

TABLE IV. True and estimated means, standard deviations (SD) and coverage (CI) of the 95% confidence intervals for the single locus simulations with random selection and with population stratification

Method	Parameter	Models											
		Popn strat (Null 2)				Popn strat (Null 2)				Popn strat (Alt)			
		True mean	Mean	(SD)	CI	True mean	Mean	(SD)	CI	True mean	Mean	(SD)	CI
Linear regression	Constant	-	2.29	(0.21)	-	-	(0.10)	-	-	-	(0.19)	-	
	β_1	0.00	-1.06	(0.26)	0.02	0.00	(0.15)	0.76	1.00	0.15	(0.24)	0.07	
QTDT _M	β_2	0.00	-2.06	(0.22)	0.00	0.00	(0.14)	0.00	2.00	0.34	(0.20)	0.00	
	β_1	0.00	0.01	(0.33)	0.95	0.00	(0.17)	0.98	1.00	1.01	(0.31)	0.95	
QCPC	β_2	0.00	0.01	(0.39)	0.94	0.00	(0.24)	0.97	2.00	2.01	(0.35)	0.94	
	β_1	0.00	0.01	(0.20)	0.95	0.00	(0.15)	0.95	0.25	0.68	(0.24)	0.61	
QPL	β_2	0.00	0.00	(0.21)	0.95	0.00	(0.16)	0.96	0.50	1.23	(0.27)	0.17	
	β_1	0.00	0.01	(0.20)	0.95	0.00	(0.15)	0.95	0.25	0.68	(0.24)	0.61	
QCPC α 's removed	β_2	0.00	0.00	(0.21)	0.95	0.00	(0.16)	0.96	0.50	1.23	(0.27)	0.17	
	β_1	0.00	0.00	(0.10)	0.96	0.00	(0.14)	0.95	0.25	0.14	(0.09)	0.81	
QPL α 's removed	β_2	0.00	0.00	(0.12)	0.95	0.00	(0.16)	0.96	0.50	0.29	(0.10)	0.50	
	β_1	0.00	0.08	(0.09)	0.91	0.00	(0.14)	0.96	0.25	0.20	(0.09)	0.92	
QCPC α 's replaced by offspring genotype (ξ)	β_2	0.00	0.01	(0.11)	0.97	0.00	(0.14)	0.96	0.50	0.30	(0.09)	0.50	
	β_1	0.00	0.00	(0.19)	0.96	0.00	(0.14)	0.95	0.25	0.62	(0.22)	0.66	
QPL α 's replaced by offspring genotype (ξ)	β_2	0.00	-0.01	(0.20)	0.96	0.00	(0.16)	0.95	0.50	1.11	(0.23)	0.25	
	β_1	0.00	-0.03	(0.19)	0.95	0.00	(0.14)	0.95	0.25	0.56	(0.22)	0.79	
	β_2	0.00	-0.07	(0.19)	0.96	0.00	(0.15)	0.95	0.50	1.00	(0.22)	0.46	

The simulation parameters are as shown in Table 1 ([online](#)).

stratification for the first null model are slightly larger than expected, across all three sampling schemes. Similarly, the type I error is inflated for the second null model under the two-tailed sampling scheme. The remainder of the methods have type I errors close to the nominal values. Therefore, the retrospective methods (QCPG and QPL) with the correct nuisance parameters or with the nuisance parameters replaced by the offspring genotype can be used (and indeed have high power) to test for a genetic effect, even under circumstances where they produce biased estimates.

MULTI-LOCUS HAPLOTYPES

Simulations were carried out to investigate the effect of the nuisance parameters (intended to account for population stratification) when the methods are extended to multi-locus haplotypes. Note that, as originally proposed, the QTDT_M method (and simple linear regression) only apply to single loci: to extend these methods to multi-locus haplotypes it is necessary to first infer the child's (and if necessary, the parents') haplotypes given the observed genotype data, as is done in the first stage of the CPG and QCPG methods [Cordell et al., 2004]. The resulting haplotype variables may then be entered as predictor variables into equation (1).

Tables V and VI show the results of simulations in which the offspring quantitative trait was influenced by genotype at two linked diallelic markers assumed to be in moderate LD. The four possible haplotypes, 1-1, 1-2, 2-1 and 2-2, had haplotype frequencies and haplotype means as shown in Table 7 (online). Additive effects of haplotypes were assumed so that for each trio, the offspring's quantitative trait was drawn from a normal distribution whose mean was the sum of the two haplotype means. For each simulation, 1,000 replicates of data were generated, each replicate consisting of 1,000 parent-offspring trios with random selection or 1,000 trios selected from 10,000 in either one-tail (top 10%) or two-tailed (top and bottom 5%) sampling from the extremes of the offspring trait distribution.

Three methods were considered, simple linear regression, QTDT_M and the retrospective method QCPG. The QPL method was not considered as it is so closely related to the QCPG. Simple linear regression does not have any additional parameters to account for population stratification. The QTDT_M and QCPG methods, however, have a significant number of nuisance parameters when the methods are extended to multi-locus haplotypes. For example, for QTDT_M, the number of possible mating types (assuming parental mating symmetry) is 55, a large increase from the 6 in the single-locus case.

Therefore, in addition to considering the models in which the 'correct' nuisance parameters are used, we considered models in which the number of nuisance parameters were reduced. For the QTDT_M we considered either including in the model the single-locus mating-type parameters for each locus, or including maternal and paternal genotype (rather than mating-type) parameters. For the QCPG method, replacing the nuisance parameters by the offspring genotype (g) was considered.

Table V shows the results for the case with no population stratification. Also shown is the number of replicates that converged from the original 1,000. Convergence problems were probably a small sample size problem, due to the large numbers of parameters to estimate in the models. Under the null, all the methods produced unbiased parameter estimates, regardless of selection scheme. Under the alternative with no selection, the prospective methods (simple linear regression and QTDT_M with the different sets of α parameters) gave unbiased parameter estimates. The retrospective QCPG method shows some small-sample bias in the estimates when the full set of 'correct' α' parameters was used and similar bias when the α' parameters were replaced by the offspring genotype g . This bias disappeared when 10,000 trios (as opposed to 1,000) were used (data not shown). Under the alternative with selection, only the retrospective QCPG method when all the 'correct' α' parameters are used, or when the α' are replaced by the offspring genotype, gave estimates close to the true mean.

The sensitivity of the estimates to the way the nuisance parameters are modelled is most pronounced under population stratification as shown in Table VI. Under the null with random selection, the simple linear regression method produces biased estimates, as does the QTDT_M method in which the correct α' s are replaced by those that would be generated by considering the loci individually. The remaining methods, QTDT_M with the full set of α parameters or with parental genotype parameters, and the QCPG methods with the different sets of nuisance parameters, all have unbiased estimates. Under the null with selection from the upper tail of the offspring trait distribution, all of the methods produced unbiased parameter estimates. For the two-tailed sampling scheme, linear regression showed the expected bias in parameter estimates but QTDT_M with the correct α' s, QTDT_M with parental genotypes and the retrospective methods (with the different sets of α' parameters) produced unbiased parameter estimates. Under the alternative with random selection only the prospective QTDT_M method (with the different sets of α' s) produced unbiased estimates: as explained in Appendices A and C, the nuisance

TABLE V. Mean estimates and standard deviations of simulated two-locus haplotype effects using the simple linear regression, QTDT_M (with different sets of nuisance parameters) and the QCPCG (with different sets of nuisance parameters)

Sampling scheme No. of replicates	Method	Parameter	Null model					Alternative model				
			Random 973	Top 980	Top+bottom 982	True mean	Mean (SD)	Random 978	Top 132	Top+bottom 292	Mean (SD)	Mean (SD)
Linear regression		β_2	0.00 (0.08)	0.00 (0.03)	-0.01 (0.17)	1.00	1.00 (0.08)	0.21 (0.06)	2.18 (0.19)			
		β_3	0.00 (0.07)	0.00 (0.03)	0.00 (0.15)	2.00	2.00 (0.07)	0.63 (0.05)	3.52 (0.05)			
		β_4	0.00 (0.07)	0.00 (0.03)	0.00 (0.14)	3.00	3.00 (0.07)	1.17 (0.06)	4.08 (0.03)			
QTDT _M all α s		β_2	0.00 (0.11)	0.00 (0.05)	0.00 (0.27)	1.00	1.00 (0.11)	0.16 (0.15)	1.70 (0.28)			
		β_3	0.00 (0.10)	0.00 (0.04)	0.00 (0.22)	2.00	2.00 (0.10)	0.59 (0.12)	3.52 (0.14)			
		β_4	0.00 (0.10)	0.00 (0.04)	0.00 (0.20)	3.00	3.00 (0.10)	1.08 (0.16)	4.50 (0.14)			
QTDT _M α s replaced by parental genotypes		β_2	0.00 (0.11)	0.00 (0.05)	0.00 (0.26)	1.00	1.00 (0.11)	0.15 (0.14)	2.21 (0.22)			
		β_3	0.00 (0.10)	0.00 (0.04)	0.00 (0.22)	2.00	2.00 (0.10)	0.57 (0.12)	3.56 (0.08)			
		β_4	0.00 (0.10)	0.00 (0.04)	0.00 (0.19)	3.00	3.00 (0.10)	1.10 (0.14)	4.12 (0.06)			
QTDT _M α s at both loci		β_2	0.00 (0.10)	0.00 (0.04)	0.00 (0.22)	1.00	1.00 (0.10)	0.14 (0.10)	1.79 (0.21)			
		β_3	-0.01 (0.09)	0.00 (0.04)	0.00 (0.19)	2.00	1.99 (0.09)	0.60 (0.10)	3.78 (0.10)			
		β_4	0.00 (0.09)	0.00 (0.04)	0.00 (0.18)	3.00	3.00 (0.09)	1.12 (0.13)	4.34 (0.10)			
QCPCG all α 's		β'_2	0.00 (0.13)	0.01 (0.31)	0.00 (0.06)	1.00	1.11 (0.18)	0.68 (0.90)	1.36 (0.96)			
		β'_3	-0.01 (0.11)	-0.01 (0.25)	0.00 (0.05)	2.00	2.23 (0.22)	1.76 (0.83)	2.69 (1.43)			
		β'_4	0.00 (0.11)	0.00 (0.25)	0.00 (0.05)	3.00	3.44 (0.36)	2.80 (0.97)	3.83 (1.56)			
QCPCG α 's replaced by ofspring genotype (g)		β'_2	0.00 (0.12)	0.00 (0.28)	0.00 (0.06)	1.00	1.06 (0.14)	0.86 (0.76)	1.32 (0.88)			
		β'_3	0.00 (0.10)	0.00 (0.24)	0.00 (0.05)	2.00	2.12 (0.19)	1.86 (0.67)	2.60 (1.10)			
		β'_4	0.00 (0.10)	0.00 (0.23)	0.00 (0.04)	3.00	3.19 (0.27)	2.87 (0.76)	3.67 (1.17)			

Simulated with no population stratification.

TABLE VI. Mean estimates and standard deviations of simulated two-locus haplotype effects using the simple linear regression, QTD_T_M (with different sets of nuisance parameters) and the QCPCG (with different sets of nuisance parameters)

Sampling scheme No. of replicates	Method	Parameter	Null model						Alternative model					
			Random 995		Top 941		Top+bottom 995		Random 996		Top 82		Top+bottom 502	
			Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Linear regression		β_2	0.00	3.30 (0.28)	0.00	0.04 (0.04)	4.14 (0.39)	1.00	4.30 (0.28)	0.26 (0.14)	0.96 (0.32)			
		β_3	0.00	2.07 (0.37)	0.00	0.05 (0.05)	2.06 (0.41)	2.00	4.07 (0.37)	0.31 (0.10)	6.74 (0.64)			
		β_4	0.00	4.07 (0.16)	0.00	0.03 (0.03)	6.22 (0.17)	3.00	7.07 (0.15)	0.50 (0.09)	9.21 (0.02)			
		β_2	0.00	-0.01 (0.42)	0.00	0.06 (0.06)	0.00 (0.51)	1.00	0.99 (0.42)	0.23 (0.19)	0.90 (0.45)			
QTD _T _M all α s		β_3	0.00	-0.01 (0.42)	0.00	0.07 (0.07)	0.00 (0.44)	2.00	1.99 (0.42)	0.30 (0.14)	5.73 (0.93)			
		β_4	0.00	-0.01 (0.33)	0.00	0.04 (0.04)	0.00 (0.41)	3.00	2.99 (0.33)	0.50 (0.13)	8.06 (0.50)			
		β_2	0.00	-0.01 (0.43)	0.00	0.06 (0.06)	-0.01 (0.51)	1.00	0.99 (0.43)	0.13 (0.23)	0.84 (0.39)			
		β_3	0.00	-0.01 (0.44)	0.00	0.06 (0.06)	0.00 (0.45)	2.00	1.99 (0.44)	0.17 (0.23)	6.48 (0.74)			
QTD _T _M α s replaced by parental genotypes		β_3	0.00	-0.01 (0.33)	0.00	0.04 (0.04)	0.00 (0.42)	3.00	2.99 (0.33)	0.37 (0.23)	8.79 (0.16)			
		β_4	0.00	-0.01 (0.37)	0.00	0.05 (0.05)	-0.01 (0.47)	1.00	1.20 (0.37)	0.24 (0.16)	0.66 (0.36)			
		β_2	0.00	0.20 (0.35)	0.00	0.05 (0.05)	0.00 (0.41)	2.00	2.23 (0.35)	0.31 (0.11)	6.45 (0.73)			
		β_3	0.00	0.04 (0.32)	0.00	0.04 (0.04)	0.00 (0.41)	3.00	3.03 (0.32)	0.51 (0.10)	8.53 (0.26)			
QTD _T _M α s at both loci		β'_2	0.00	0.00 (0.05)	0.00	0.36 (0.36)	0.00 (0.03)	2.00	-0.03 (0.08)	8.83 (12.9)	1.07 (0.55)			
		β'_3	0.00	0.00 (0.05)	0.00	0.23 (0.23)	0.00 (0.02)	3.00	0.20 (0.09)	9.59 (12.7)	3.30 (1.73)			
		β'_4	0.00	0.00 (0.03)	0.00	0.29 (0.29)	0.00 (0.02)	1.00	0.44 (0.10)	10.51 (12.7)	4.13 (1.77)			
		β'_2	0.00	0.00 (0.03)	0.01 (0.32)	0.01 (0.32)	0.00 (0.02)	2.00	0.08 (0.04)	7.64 (9.15)	1.05 (0.50)			
QCPCG all α 's		β'_3	0.00	0.00 (0.03)	0.01 (0.32)	0.01 (0.32)	0.00 (0.02)	2.00	0.14 (0.04)	8.41 (9.01)	3.46 (1.85)			
		β'_4	0.00	0.00 (0.03)	0.00 (0.22)	0.00 (0.22)	0.00 (0.02)	3.00	0.31 (0.05)	9.31 (9.03)	4.30 (1.92)			
		β'_2	0.00	0.00 (0.03)	0.00 (0.02)	0.00 (0.02)	0.00 (0.02)	1.00	0.08 (0.04)	7.64 (9.15)	1.05 (0.50)			
		β'_3	0.00	0.00 (0.03)	0.01 (0.32)	0.01 (0.32)	0.00 (0.02)	2.00	0.14 (0.04)	8.41 (9.01)	3.46 (1.85)			
QCPCG α 's replaced by offspring genotype (ζ)		β'_4	0.00	0.00 (0.03)	0.00 (0.22)	0.00 (0.22)	0.00 (0.02)	3.00	0.31 (0.05)	9.31 (9.03)	4.30 (1.92)			

Simulated with population stratification.

parameters for the QPL and QCPG will not be correctly estimated under population stratification, except under the null. Under the alternative with selection, all the methods gave biased estimates as found in the single-locus case.

We also investigated the QTDT_M and QCPG methods with a single replicate of data generated under a three-locus haplotype model (data not shown). Results were broadly similar to the two-locus haplotype results, except that the QCPG method required a very large number (50,000) trios to produce unbiased estimates, while QTDT_M generally achieved convergence and unbiased parameter estimation with only 1,000 trios.

STEPWISE PROCEDURE

A stepwise procedure (results not shown), as used by Cordell et al. [2004] for disease traits, was used to compare the prospective QTDT_M (using the full set of nuisance parameters) with the QCPG method with the 'correct' α parameters replaced by the offspring genotype, under models with and without population stratification and selection. In general, the pattern of results in terms of power and type 1 error was as expected, with the QTDT_M method being the more powerful in general. Under population stratification we found that the Type I errors were slightly too large for the QTDT_M method under random sampling, consistent with the results observed (Table 6 (**online**)) in the single locus simulations. Additional simulations (data not shown) indicated that this problem could be solved by use of Wald tests incorporating robust 'information sandwich' variance estimates [Huber, 1967], rather than likelihood ratio tests or Wald tests with the usual variance estimate (which equals minus the inverse of the Hessian matrix). We also investigated the power and type 1 error of the stepwise approach when applied to non-normally distributed traits and found that both QTDT_M and QCPG appear to be suitable for the analysis of traits that deviate slightly from normality. Neither method was found to be suitable for the analysis of very nonnormally distributed traits, although it is worth noting that the prospective QTDT_M method could easily be extended to enable the analysis of nonnormal traits by use of robust regression, a generalised linear model (GLM), or by assuming a variance-mean relationship, according to the departure from normality.

MATERNAL GENOTYPE AND PARENT-OF-ORIGIN EFFECTS

The previous single-locus simulations were modified to include maternal genotype and parent-of-origin effects. For each replicate, 1,000 trios were generated in which the offspring's quantitative trait

was influenced by its own genotype, and by either the mother's genotype, or whether the offspring received a variant allele from the mother, or both. Under the alternative, 100 replicates of data were generated. Under the null (no maternal genotype effects or no parent-of-origin effects) 1,000 replicates of data were generated. The QCEPG and QTDT_M methods were implemented in Stata. For Kistner and Weinberg's approach [Kistner et al., 2006], the SAS macro provided at <http://dir.niehs.nih.gov/dirbb/weinbergfiles/qpl.htm> was used. The expected effect estimates for QCEPG and QTDT_M should be the same, since the traits were simulated to have unit variance. The offspring reference category was chosen to be the 1/1 genotype, and β_1 and β_2 are the estimated effects for the 1/2 (2/1) and 2/2 genotypes respectively. Similarly, maternal genotype effects are denoted as β_{m1} and β_{m2} , and parent-of-origin effects by β_j . However, the expected estimates for Kistner and Weinberg's method should be slightly different. In their method the reference category for the offspring genotype effects is the heterozygous genotype, rather than the homozygous (1/1) genotype. The maternal effects, denoted by δ_{01} and δ_{12} , compare the difference in quantitative trait for a mother with 1 variant allele to a mother with 0 variant alleles, and a mother with 2 variant alleles to a mother with 1 variant allele respectively (while in the QCEPG method, both comparisons are made with the homozygous 1/1 genotype category). The estimates for the parent-of-origin effects (λ_1) in the QPL represent the log odds that a heterozygous child inherits a maternal copy of the variant allele instead of a paternal copy, per unit increase in trait value. Although based only on heterozygous offspring, these parameters are expected to be the same as for the QCEPG and QTDT_M methods.

Tables 8 and 9 (**online**) show the true effects, the estimated means and standard deviations. All three methods produce reasonable estimates under the null. The results for the prospective QTDT_M method show the least bias. Under the alternative, the retrospective methods show a bias when parent-of-origin effects are present. The QPL appears to produce parent-of-origin effects of approximately 0.5, when they would have been expected to be 1. This may be due to some unrecognised difference in the parameterizations: the QCEPG uses the original parent-of-origin parameterization of Weinberg et al. [1998], whereas the QPL uses a parameterization closer to the alternative parameterization suggested by Weinberg [1999]. Table 10 (**online**) shows the powers and type I errors. The type I errors for the QPL and QTDT_M methods seem reasonable. However, the type I error for the QCEPG method when testing maternal genotype effects is very large, although this appears to be a small-sample issue as

it improved in simulations with a larger number of trios (data not shown). Overall, the extension of the QPL method had the highest power to detect either a maternal genotype or parent-of-origin effect.

DISCUSSION

In this paper, we have extended the case/pseudocontrol association approach for dichotomous phenotypes [Cordell and Clayton, 2002] to perform association analysis with quantitative traits. This approach is very similar to the QPL approach proposed by Kistner and Weinberg [2004], but uses a slightly more intuitive parameterization and extends more naturally to allow analysis of multiallelic markers, multiple linked loci, multiple unlinked regions, parent-of-origin or maternal genotype effects, gene-gene and gene-environment interactions, using the same formulation as Cordell et al. [2004]. We compared this approach to a prospective approach, the QTDT_M and also extended the QTDT_M to allow analysis of multiple linked loci (including multi-locus haplotypes), parent-of-origin or maternal genotype effects. Other extensions to the QTDT_M follow naturally.

All the methods incorporate nuisance parameters intended to account for population stratification. When considering multi-locus haplotypes, the number of nuisance parameters can dramatically increase, and so it is important to find ways to reduce the number of nuisance parameters. It was found that replacing the nuisance parameters by the offspring genotype in the retrospective methods worked almost as well as the full model, and replacing the nuisance parameters by parental genotypes worked well for the QTDT_M. In our simulations, it was assumed that both parents came from the same sub-population. If, in fact, matings occurred between individuals from different sub-populations, one might not expect these approximations to work as well as fitting the full set of nuisance parameters.

Although the retrospective approaches had some advantages with regard to estimation of parameters under selected sampling, in general we found the prospective QTDT_M to be the most efficient approach, requiring smaller sample sizes to achieve convergence and asymptotic behaviour. In addition, the parameter estimates provided by the QTDT_M have a more intuitive interpretation, corresponding to the direct genotype effects on the trait, whereas the retrospective approaches estimate parameters that are scaled by division by the unknown (although potentially estimable) trait variance. Covariates are also more easily incorporated into the QTDT_M framework, simply by adding them in as terms in the regression equation, although it would

be possible to incorporate covariates in the retrospective approaches, either by first regressing the traits on covariates of interest and performing subsequent analysis on the residuals, or by using methods such as those described by Lim et al. [2005].

The QTDT_M method was found to be the only method suitable for estimation of effects under the alternative hypothesis with population stratification (assuming random sampling). Under population stratification, it was necessary to use robust 'information sandwich' variance estimates to achieve correct type 1 errors and confidence interval coverage with the QTDT_M. This is possibly because the parental mating-type stratification parameters act as a surrogate for population membership in the sense that they soak up the mean level of bias induced by population stratification, but do they not fully account for population membership, so that the distribution of trait within parental mating-type classes violates the assumption of normality, even if this assumption holds within each sub-population.

The analyses described here assumed availability of a dataset consisting of parent-offspring trios, with no missing genotype data. A natural extension of the methods proposed here would be to consider analysis of large extended pedigrees and/or missing genotype data. The QPL has previously been extended to allow analysis of multiple siblings and missing parents [Kistner and Weinberg, 2005] while an approach asymptotically equivalent to QTDT_M, namely the HQTDT of Abecasis et al. [2000], has been extended to apply to pedigrees of arbitrary structure [Abecasis et al., 2002]. However, these approaches focus on testing rather than estimation of effects and apply only to a single locus at a time. A natural way to extend the QCEPG and QTDT_M approaches developed here for analysis of general pedigrees would be to perform tests using Wald tests and incorporate robust 'information sandwich' variance estimates that cluster observations according to pedigree [Huber, 1967]. An alternative approach would be to use a random-effects modelling framework [Xu and Shete, 2006]. With regards to missing genotype data, methods that sample or average over the possible genotype configurations consistent with the observed genotype data, in the correct proportions [Cordell, 2006], could be considered. Investigation of these approaches and their behaviour under complex disease models, in the presence of population stratification, will form the basis of future work.

ACKNOWLEDGMENTS

Support for this work was provided by the Wellcome Trust (Grant references 074524

and 068612). We thank Jenefer Blackwell, Joanna Biernacka and Jeff O'Connell for useful discussions, and to Joanna Howson and Hin-Tak Leung for technical assistance.

REFERENCES

- Abecasis GR, Cardon LR, Cookson WOC. 2000. A general test of association for quantitative traits in nuclear families. *Am J Hum Genet* 66:279–292.
- Abecasis GR, Cherny SS, Cookson WO, Cardon LR. 2002. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 30:97–101.
- Allison DB. 1997. Transmission-disequilibrium tests for quantitative traits. *Am J Hum Genet* 60:676–690.
- Cordell HJ, Barratt BJ, Clayton DG. 2004. Case/pseudocontrol analysis in genetic association studies: a unified framework for detection of genotype and haplotype associations, gene-gene and gene-environment interactions, and parent-of-origin effects. *Genet Epidemiol* 26:167–185.
- Cordell HJ, Clayton DG. 2002. A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. *Am J Hum Genet* 70:124–141.
- Cordell HJ. 2006. Estimation and testing of genotype and haplotype effects in case/control studies: comparison of weighted regression and multiple imputation procedures. *Genet Epidemiol* 30:259–275.
- Fulker DW, Cherny SS, Sham PC, Hewitt JK. 1999. Combined linkage and association sib-pair analysis for quantitative traits. *Am J Hum Genet* 64:259–267.
- Gauderman WJ. 2003. Candidate gene association analysis for a quantitative trait, using parent-offspring trios. *Genet Epidemiol* 25:327–338.
- Huber PJ. 1967. The behaviour of maximum likelihood estimates under nonstandard conditions. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 221–233.
- Kistner EO, Infante-Rivard C, Weinberg CR. 2006. A method for using incomplete triads to test maternally mediated genetic effects and parent-of-origin effects in relation to a quantitative trait. *Am J Epidemiol* 163:255–261.
- Kistner EO, Weinberg CR. 2004. Method for using complete and incomplete trios to identify genes related to a quantitative trait. *Genet Epidemiol* 27:33–42.
- Kistner EO, Weinberg CR. 2005. A method for identifying genes related to a quantitative trait, incorporating multiple siblings and missing parents. *Genet Epidemiol* 29:155–165.
- Laird NM, Horvath S, Xu X. 2000. Implementing a unified approach to family-based tests of association. *Genet Epidemiol* 19 (suppl 1):S36–S42.
- Lange C, DeMeo DL, Laird NM. 2002. Power and design considerations for a general class of family-based association tests: quantitative traits. *Am J Hum Genet* 71:1330–1341.
- Lim S, Beyenne J, Greenwood CMT. 2005. Continuous covariates in genetic association studies of case-parent triads: gene-environment interaction effects, population stratification, and power analysis. *Statist Appl Genetics Mol Biol* 4:1–25.
- Liu Y, Tritchler D, Bull SB. 2002. A unified framework for transmission-disequilibrium test analysis of discrete and continuous traits. *Genet Epidemiol* 22:26–40.
- Lunetta K, Faraone S, Biederman J, Laird N. 2000. Family-based tests of association and linkage that use unaffected sibs, covariates, and interactions. *Am J Hum Genet* 66:605–614.
- Schaid DJ, Sommer SS. 1993. Genotype relative risks: methods for design and analysis of candidate-gene association studies. *Am J Hum Genet* 53:1114–1126.
- Schaid DJ. 1996. General score tests for associations of genetic markers with disease using cases and their parents. *Genet Epidemiol* 13:423–449.
- Spielman RS, McGinnis RE, Ewens WJ. 1993. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52:506–1993.
- Weinberg CR, Wilcox AJ, Lie RT. 1998. A log-linear approach to case-parent-triad data: assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. *Am J Hum Genet* 62:969–978.
- Weinberg CR. 1999. Methods for detection of parent-of-origin effects in genetic studies of case-parents triads. *Am J Hum Genet* 65:229–235.
- Whittaker JC, Gharani N, Hindmarsh P, McCarthy MI. 2003. Estimation and testing of parent-of-origin effects for quantitative traits. *Am J Hum Genet* 72:1035–1039.
- Xu H, Shete S. 2006. Mixed-effects logistic approach for association following linkage scan for complex disorders. *Hum Hered*, in press.
- Yang Q, Rabinowitz D, Isasi C, Shea S. 2000. Adjusting for confounding due to population admixture when estimating the effect of candidate genes on quantitative traits. *Hum Hered* 50:227–233.

APPENDIX A

EXPRESSION OF QCPG LIKELIHOOD

The retrospective QCPG likelihood is parameterized in terms of parameters of interest (offspring genotype effects, denoted β') and several nuisance parameters (denoted α') as shown in Table II. Here we express the β' and α' parameters in the QCPG likelihood in terms of various parameters (denoted α and β) in a prospective model for trait given genotype. We do not propose to reparameterize the QCPG likelihood of Table II in terms of these prospective parameters before maximization. Rather, we continue to freely estimate the β' and α' parameters when we maximise the QCPG likelihood. However, we use the relationship between the

retrospective and prospective parameters to inform our understanding and interpretation of the retrospective parameters, since the α and β parameters in the prospective model are generally more intuitively interpretable than those on the retrospective scale.

For a normally distributed quantitative trait Y with mean μ and variance σ^2 , the probability distribution function of the observed trait y_i is given by

$$f(y_i; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \mu)^2}{2\sigma^2}\right) \tag{7}$$

For a trio with parental mating type M and offspring genotype g_i , μ equals the mean of the prospective QTDT_M approach from equation (1), that is $\mu = \alpha_M + \beta_{g_i}$. Equation (7) becomes

$$f(y_i; \alpha_M, \beta_{g_i}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - (\alpha_M + \beta_{g_i}))^2}{2\sigma^2}\right) \tag{8}$$

From equation (3), the contribution of a trio to the likelihood can be expressed as

$$P(g_i | g_{im}, g_{if}, y_i) = \frac{P(y_i | g_i) P(g_i | g_{im}, g_{if})}{\sum_{g_i^* \in S'_M} P(y_i | g_i^*) P(g_i^* | g_{im}, g_{if})} \tag{9}$$

So, from equations (8) and (9), the contribution of a trio to the likelihood for a normally distributed trait with mean $\alpha_M + \beta_{g_i}$ and variance σ^2 is given by

$$\begin{aligned} P(g_i | g_{im}, g_{if}, y_i) &= \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - (\alpha_M + \beta_{g_i}))^2}{2\sigma^2}\right) P(g_i | g_{im}, g_{if})}{\sum_{g_i^* \in S'_M} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - (\alpha_M + \beta_{g_i^*}))^2}{2\sigma^2}\right) P(g_i^* | g_{im}, g_{if})} \\ &= \frac{\exp\left(-\frac{(y_i - (\alpha_M + \beta_{g_i}))^2}{2\sigma^2}\right) P(g_i | g_{im}, g_{if})}{\sum_{g_i^* \in S'_M} \exp\left(-\frac{(y_i - (\alpha_M + \beta_{g_i^*}))^2}{2\sigma^2}\right) P(g_i^* | g_{im}, g_{if})} \\ &= \frac{\exp\left(-\frac{y_i^2}{2\sigma^2} + \frac{2y_i(\alpha_M + \beta_{g_i})}{2\sigma^2} - \frac{(\alpha_M + \beta_{g_i})^2}{2\sigma^2}\right) P(g_i | g_{im}, g_{if})}{\sum_{g_i^* \in S'_M} \exp\left(-\frac{y_i^2}{2\sigma^2} + \frac{2y_i(\alpha_M + \beta_{g_i^*})}{2\sigma^2} - \frac{(\alpha_M + \beta_{g_i^*})^2}{2\sigma^2}\right) P(g_i^* | g_{im}, g_{if})} \end{aligned} \tag{10}$$

In each offspring's contribution to the likelihood, the quantitative trait, y_i , its variance, σ^2 , and the parental mating type parameter α_M are the same for the offspring (the 'case') and the pseudocontrols. Therefore, cancelling terms from the numerator and denominator,

$$\begin{aligned} P(g_i | g_{im}, g_{if}, y_i) &= \frac{\exp\left(\frac{y_i \beta_{g_i}}{\sigma^2} - \frac{(\alpha_M + \beta_{g_i})^2}{2\sigma^2}\right) P(g_i | g_{im}, g_{if})}{\sum_{g_i^* \in S'_M} \exp\left(\frac{y_i \beta_{g_i^*}}{\sigma^2} - \frac{(\alpha_M + \beta_{g_i^*})^2}{2\sigma^2}\right) P(g_i^* | g_{im}, g_{if})} \\ &= \frac{\exp\left(\frac{\beta_{g_i}}{\sigma^2} y_i - \frac{\alpha_M^2 + 2\alpha_M \beta_{g_i} + \beta_{g_i}^2}{2\sigma^2}\right) P(g_i | g_{im}, g_{if})}{\sum_{g_i^* \in S'_M} \exp\left(\frac{\beta_{g_i^*}}{\sigma^2} y_i - \frac{\alpha_M^2 + 2\alpha_M \beta_{g_i^*} + \beta_{g_i^*}^2}{2\sigma^2}\right) P(g_i^* | g_{im}, g_{if})} \\ &= \frac{\exp\left(\frac{\beta_{g_i}}{\sigma^2} y_i - \frac{\beta_{g_i}(2\alpha_M + \beta_{g_i})}{2\sigma^2}\right) P(g_i | g_{im}, g_{if})}{\sum_{g_i^* \in S'_M} \exp\left(\frac{\beta_{g_i^*}}{\sigma^2} y_i - \frac{\beta_{g_i^*}(2\alpha_M + \beta_{g_i^*})}{2\sigma^2}\right) P(g_i^* | g_{im}, g_{if})} \end{aligned}$$

which can be written in the following form:

$$P(g_i|g_{im}, g_{if}, y_i) = \frac{\exp\left(\left[\frac{\beta_{g_i^*}}{\sigma^2}y_i\right] + \left[-\frac{\beta_{g_i^*}(2\alpha_M + \beta_{g_i^*})}{2\sigma^2} + \ln(P(g_i|g_{im}, g_{if}))\right]\right)}{\sum_{g_i^* \in S_M} \exp\left(\left[\frac{\beta_{g_i^*}}{\sigma^2}y_i\right] + \left[-\frac{\beta_{g_i^*}(2\alpha_M + \beta_{g_i^*})}{2\sigma^2} + \ln(P(g_i^*|g_{im}, g_{if}))\right]\right)} \tag{11}$$

This is in the general form of equation (4), where

$$\beta'_{g_i^*} = \frac{\beta_{g_i^*}}{\sigma^2} \text{ and } \alpha'_{Mg_i^*} = -\frac{\beta_{g_i^*}(2\alpha_M + \beta_{g_i^*})}{2\sigma^2} + \ln(P(g_i^*|g_{im}, g_{if}))$$

Thus the parameters from the retrospective QCPG formulation (equation (4)) may be interpreted as follows. The term $\beta'_{g_i^*} = \frac{\beta_{g_i^*}}{\sigma^2}$ is the original genotype effect of genotype g_i^* on the prospective scale, divided by the trait variance. (Hence, given parameter estimates $\beta'_{g_i^*}$ from fitting the QCPG model, the genotype effects on the prospective scale could be obtained by multiplying the estimate of $\beta'_{g_i^*}$ by the corresponding trait variance, if it were known or estimable.) The terms $\alpha'_{Mg_i^*} = -\beta_{g_i^*}(2\alpha_M + \beta_{g_i^*})/2\sigma^2 + \ln(P(g_i^*|g_{im}, g_{if}))$ correspond to nuisance parameters that allow the model to account for non-Mendelianism and population stratification, as proposed by Kistner and Weinberg [2004]. As discussed in the text, not all of the β' and α' are estimable, so restrictions must be made such as setting some of these equal to zero (equivalent to choosing a reference genotype category to which the other genotype effects are compared). This complicates the interpretation of the nuisance parameters. Suppose that all parameters were in fact estimable. Suppose further that there is no non-Mendelianism, so that $P(g_i^*|g_{im}, g_{if}) = 0.25$ for all offspring genotypes consistent with the parental genotypes. In that case, we could write out in full the relationships between the retrospective and prospective parameters implied by equation (11):

$$\begin{aligned} \beta'_0 &= \frac{\beta_0}{\sigma^2}, \beta'_1 = \frac{\beta_1}{\sigma^2}, \beta'_2 = \frac{\beta_2}{\sigma^2} \\ \alpha'_{010} &= -\frac{\beta_0(2\alpha_{01} + \beta_0)}{2\sigma^2} + \ln(0.25) \\ \alpha'_{011} &= -\frac{\beta_1(2\alpha_{01} + \beta_1)}{2\sigma^2} + \ln(0.25) \\ \alpha'_{110} &= -\frac{\beta_0(2\alpha_{11} + \beta_0)}{2\sigma^2} + \ln(0.25) \\ \alpha'_{111} &= -\frac{\beta_1(2\alpha_{11} + \beta_1)}{2\sigma^2} + \ln(0.25) \\ \alpha'_{112} &= -\frac{\beta_2(2\alpha_{11} + \beta_2)}{2\sigma^2} + \ln(0.25) \\ \alpha'_{121} &= -\frac{\beta_1(2\alpha_{12} + \beta_1)}{2\sigma^2} + \ln(0.25) \\ \alpha'_{122} &= -\frac{\beta_2(2\alpha_{12} + \beta_2)}{2\sigma^2} + \ln(0.25) \end{aligned}$$

In practice, not all the parameters are estimable and we set $\beta'_1, \alpha'_{011}, \alpha'_{111}$ and α'_{121} (the reference parameters) equal to zero. Assuming that on the prospective scale we also set $\beta_1 = 0$, and rearranging the expressions above, the relationships become

$$\begin{aligned} \beta'_0 &= \frac{\beta_0}{\sigma^2}, \beta'_1 = 0, \beta'_2 = \frac{\beta_2}{\sigma^2} \\ \alpha'_{010} &= \alpha'_{011} + \frac{\beta_1(2\alpha_{01} + \beta_1)}{2\sigma^2} - \frac{\beta_0(2\alpha_{01} + \beta_0)}{2\sigma^2} = -\frac{\beta_0(2\alpha_{01} + \beta_0)}{2\sigma^2} \\ \alpha'_{110} &= \alpha'_{111} + \frac{\beta_1(2\alpha_{11} + \beta_1)}{2\sigma^2} - \frac{\beta_0(2\alpha_{11} + \beta_0)}{2\sigma^2} = -\frac{\beta_0(2\alpha_{11} + \beta_0)}{2\sigma^2} \\ \alpha'_{112} &= \alpha'_{111} + \frac{\beta_1(2\alpha_{11} + \beta_1)}{2\sigma^2} - \frac{\beta_2(2\alpha_{11} + \beta_2)}{2\sigma^2} = -\frac{\beta_2(2\alpha_{11} + \beta_2)}{2\sigma^2} \\ \alpha'_{122} &= \alpha'_{121} + \frac{\beta_1(2\alpha_{12} + \beta_1)}{2\sigma^2} - \frac{\beta_2(2\alpha_{12} + \beta_2)}{2\sigma^2} = -\frac{\beta_2(2\alpha_{12} + \beta_2)}{2\sigma^2} \end{aligned}$$

Under the null hypothesis, the prospective genotype effects β_0 and β_2 equal zero, and so the true values of the retrospective parameters (both the genotype parameters of interest and the nuisance parameters) will also equal zero.

A similar argument applies if, instead of setting β_1 and β'_1 to zero, we set β_0 and β'_0 to zero, so that the genotype parameters are all calculated relative to the homozygous 1/1 genotype. In this case, we still find that under the null hypothesis that β_1 and β_2 equal zero, the true values of the retrospective parameters (both the genotype parameters of interest and the nuisance parameters) equal zero (data not shown).

The above derivation for the nuisance parameters only works because $P(g_i^* | g_{im}, g_{if}) = 0.25$ in all cases and so is subtracted out when we express α'_{010} , α'_{110} , α'_{112} and α'_{122} in terms of α_{011} , α'_{111} and α'_{121} . If, in fact, there was non-Mendelianism, we would have instead that

$$\begin{aligned}\alpha'_{jkl} &= -\frac{\beta_l(2\alpha_{jkl} + \beta_l)}{2\sigma^2} + \ln(P(g_i^* = l | \{g_{im}, g_{if}\} = \{j, k\})) \\ &\quad - \ln(P(g_i^* = 1 | \{g_{im}, g_{if}\} = \{j, k\})) \\ &= -\frac{\beta_l(2\alpha_{jkl} + \beta_l)}{2\sigma^2} + \psi_{jkl}, \text{ say.}\end{aligned}$$

In this case the term ψ_{jkl} will be nonzero, but fitting α'_{jkl} allows the model to account for this, so that even under the null hypothesis where β_0 and β_2 equal zero, the nuisance parameters α'_{jkl} will not equal zero.

The effect of population stratification is modelled on the prospective scale via $\mu = \alpha_M + \beta_{g_i}$ which allows the mean trait value for a child to vary according to the genotype of its parents. If there really were population stratification, the correct model would in fact be $\mu = \mu_p + \beta_{g_i}$, where μ_p denotes the mean trait value in the (unknown) sub-population p from which the child originates. If we replace α_M by μ_p everywhere above, we find that

$$\alpha'_{jkl} = -\frac{\beta_l(2\mu_p + \beta_l)}{2\sigma^2} + \psi_{jkl}$$

Thus the true value of the nuisance parameters should vary according to sub-population, which we do not allow for. However, under the null hypothesis where β_0 and β_2 equal zero, α'_{jkl} does not vary by sub-population and so we should still obtain valid inference for the β' under the null, even though the model misspecification means we may not obtain valid inference under the alternative.

APPENDIX B

EXPRESSION OF QPL LIKELIHOOD

We may use a similar argument as in Appendix A to derive the relationship between the α'' and β'' parameters in the QPL model and the prospective α and β parameters. The only difference is in the summation in the denominator of equation (11), which is over either two or three possible offspring genotype categories corresponding to the categories shown in Table I, (e.g. 1/1, 1/2 (unordered) and 2/2 for offspring of two heterozygous parents, rather than over four genotype categories 1/1, 1/2, 2/1 and 2/2). The result of this is that the relationships implied by equation (11) become

$$\begin{aligned}\beta''_0 &= \frac{\beta_0}{\sigma^2}, \beta''_1 = \frac{\beta_1}{\sigma^2}, \beta''_2 = \frac{\beta_2}{\sigma^2} \\ \alpha''_{010} &= -\frac{\beta_0(2\alpha_{01} + \beta_0)}{2\sigma^2} + \ln(0.5) \\ \alpha''_{011} &= -\frac{\beta_1(2\alpha_{01} + \beta_1)}{2\sigma^2} + \ln(0.5) \\ \alpha''_{110} &= -\frac{\beta_0(2\alpha_{11} + \beta_0)}{2\sigma^2} + \ln(0.25) \\ \alpha''_{111} &= -\frac{\beta_1(2\alpha_{11} + \beta_1)}{2\sigma^2} + \ln(0.5) \\ \alpha''_{112} &= -\frac{\beta_2(2\alpha_{11} + \beta_2)}{2\sigma^2} + \ln(0.25) \\ \alpha''_{121} &= -\frac{\beta_1(2\alpha_{12} + \beta_1)}{2\sigma^2} + \ln(0.5) \\ \alpha''_{122} &= -\frac{\beta_2(2\alpha_{12} + \beta_2)}{2\sigma^2} + \ln(0.5)\end{aligned}$$

where the terms $P(g_i^* | g_{im}, g_{if})$ differ according to offspring genotype, for offspring of two heterozygous parents. If, as in Appendix A, we set $\beta_1, \beta'_1, \alpha'_{011}, \alpha'_{111}$ and α'_{121} to zero, we obtain the following equations:

$$\begin{aligned}\beta''_0 &= \frac{\beta_0}{\sigma^2}, \beta''_1 = 0, \beta''_2 = \frac{\beta_2}{\sigma^2} \\ \alpha''_{010} &= \alpha'_{011} + \frac{\beta_1(2\alpha_{01} + \beta_1)}{2\sigma^2} - \frac{\beta_0(2\alpha_{01} + \beta_0)}{2\sigma^2} = -\frac{\beta_0(2\alpha_{01} + \beta_0)}{2\sigma^2} \\ \alpha''_{110} &= \alpha'_{111} - \ln(0.5) + \ln(0.25) + \frac{\beta_1(2\alpha_{11} + \beta_1)}{2\sigma^2} - \frac{\beta_0(2\alpha_{11} + \beta_0)}{2\sigma^2} = -\frac{\beta_0(2\alpha_{11} + \beta_0)}{2\sigma^2} + \ln(0.5) \\ \alpha''_{112} &= \alpha'_{111} - \ln(0.5) + \ln(0.25) + \frac{\beta_1(2\alpha_{11} + \beta_1)}{2\sigma^2} - \frac{\beta_2(2\alpha_{11} + \beta_2)}{2\sigma^2} = -\frac{\beta_2(2\alpha_{11} + \beta_2)}{2\sigma^2} + \ln(0.5) \\ \alpha''_{122} &= \alpha'_{121} + \frac{\beta_1(2\alpha_{12} + \beta_1)}{2\sigma^2} - \frac{\beta_2(2\alpha_{12} + \beta_2)}{2\sigma^2} = -\frac{\beta_2(2\alpha_{12} + \beta_2)}{2\sigma^2}\end{aligned}$$

Under the null hypothesis that β_0 and β_2 equal zero, the α''_{jkl} do not all equal zero, and so it is necessary in the QPL model that the α''_{jkl} be included in order to obtain correct inference, even when there is no population stratification or non-Mendelianism.

APPENDIX C

DISCUSSION OF NUISANCE PARAMETERS

A difference between the QCPG and QPL methods arises with regard to the number of nuisance parameters estimated. It can be seen in Table I that there are a total of four nuisance parameters, $\alpha_{010}, \alpha_{110}, \alpha_{112}, \alpha_{122}$. However, by fitting the model using a polytomous logistic model as proposed by Kistner and Weinberg, two additional, essentially inestimable, nuisance parameters are estimated. These are referred to in Kistner and Weinberg ([2004] p. 36) as α_{012} and α_{120} , corresponding to the situations where either one parent has no copies of the variant allele but the offspring has two copies, or one parent has two copies but the offspring has none. There is no data to estimate these parameters (they correspond to impossible events) but the computer program tries to estimate them since they are included in the model. Implementation of the QPL approach using SAS code available from <http://dir.niehs.nih.gov/dirbb/weinbergfiles/qpl.htm> (data not shown) indicates that estimation of these two inestimable α_{012} and α_{120} parameters is poor, with very large estimates and confidence intervals, although this does not appear to adversely affect estimation of the six genuinely estimable parameters.

For the QCPG and QPL methods, if there are genotype effects present, the nuisance parameters are non-zero. When population stratification exists, the true values of these nuisance parameters may differ in the different sub-populations (due to differences in α_M between subpopulations). If estimated in the combined population, the nuisance parameters will be an average of those from the different sub-populations and so will not necessarily provide accurate population-specific estimates. Hence, although the type 1 errors will be correct, we do not necessarily expect the QCPG (or QPL) method to have unbiased parameter estimates when used for estimation of effects under the alternative hypothesis with population stratification, even though they should be unbiased under the null.

When extending the QCPG method to multi-locus haplotypes, the number of nuisance parameters dramatically increases since the number of parental mating types and offspring genotypes increases. The nuisance parameters in the single-locus case are of the form $\alpha'_{Mg_i^*} = -\beta_{g_i^*}(2\alpha_M + \beta_{g_i^*})/2\sigma^2 + \ln(P(g_i^* | g_{im}, g_{if}))$ (equation (11)).

Through simulations (described in the main text), we investigate whether the parental genotype information still required in the nuisance parameters can be captured if the nuisance parameters are replaced by parameters representing the offspring genotype alone ($\gamma_{g_i^*}$), thus reducing the overall number of nuisance parameters. Equation (11) would then become

$$P(g_i | g_{im}, g_{if}, y_i) = \frac{\exp\left(\left[\frac{\beta_{g_i^*}}{\sigma^2} y_i\right] + \gamma_{g_i^*}\right)}{\sum_{g_i^* \in S'_M} \exp\left(\left[\frac{\beta_{g_i^*}}{\sigma^2} y_i\right] + \gamma_{g_i^*}\right)}$$

The hope is that the population stratification can be accounted for by offspring genotype alone, rather than by a combination of offspring and parental genotypes.

APPENDIX D

EXPRESSION OF THE QCEPG LIKELIHOOD

The rationale for expressing the likelihood equation (5) in the form of equation (6) is very similar to that used for the QCPCG approach. Previously, for the QCPCG method, the population mean of the quantitative trait Y (μ) was assumed to equal the mean of the prospective QTDT_M method such that for an offspring with genotype β_{g_i} , $\mu = \alpha_M + \beta_{g_i}$, for mating type M and offspring genotype effect β_{g_i} . To include maternal genotype or parent-of-origin effects, the mean is now assumed to depend on parental mating type and the offspring genotype effect as before, but it now also depends on the maternal genotype effect, $\beta_{g_{im}}$, and the parent-of-origin effect β_{I_m} . Therefore,

$$\mu = \alpha_M + \beta_{g_i} + \beta_{g_{im}} + \beta_{I_m}$$

Hence, the trait distribution can be expressed as a normal distribution with mean μ as above, and substituting this into equation (5) gives

$$\begin{aligned} P(g_i|g_{im}, g_{if}, y_i, \xi) &= \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - (\alpha_M + \beta_{g_i} + \beta_{g_{im}} + \beta_{I_m}))^2}{2\sigma^2}\right) P(g_i|g_{im}, g_{if})}{\sum_{g_i^*, g_{im}^*, g_{if}^* \in G_\xi} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - (\alpha_M + \beta_{g_i^*} + \beta_{g_{im}^*} + \beta_{I_m}^*))^2}{2\sigma^2}\right) P(g_i^*|g_{im}^*, g_{if}^*)} \\ &= \frac{\exp\left(-\frac{(y_i^2 - 2y_i(\alpha_M + \beta_{g_i} + \beta_{g_{im}} + \beta_{I_m}) + (\alpha_M + \beta_{g_i} + \beta_{g_{im}} + \beta_{I_m})^2)}{2\sigma^2}\right) P(g_i|g_{im}, g_{if})}{\sum_{g_i^*, g_{im}^*, g_{if}^* \in G_\xi} \exp\left(-\frac{(y_i^2 - 2y_i(\alpha_M + \beta_{g_i^*} + \beta_{g_{im}^*} + \beta_{I_m}^*) + (\alpha_M + \beta_{g_i^*} + \beta_{g_{im}^*} + \beta_{I_m}^*)^2)}{2\sigma^2}\right) P(g_i^*|g_{im}^*, g_{if}^*)} \\ &= \frac{\exp\left(\frac{y_i(\alpha_M + \beta_{g_i} + \beta_{g_{im}} + \beta_{I_m})}{\sigma^2} - \frac{(\alpha_M + \beta_{g_i} + \beta_{g_{im}} + \beta_{I_m})^2}{2\sigma^2}\right) P(g_i|g_{im}, g_{if})}{\sum_{g_i^*, g_{im}^*, g_{if}^* \in G_\xi} \exp\left(\frac{y_i(\alpha_M + \beta_{g_i^*} + \beta_{g_{im}^*} + \beta_{I_m}^*)}{\sigma^2} - \frac{(\alpha_M + \beta_{g_i^*} + \beta_{g_{im}^*} + \beta_{I_m}^*)^2}{2\sigma^2}\right) P(g_i^*|g_{im}^*, g_{if}^*)} \\ &= \frac{\exp\left(\frac{\beta_{g_i}}{\sigma^2} y_i + \frac{\beta_{g_{im}}}{\sigma^2} y_i + \frac{\beta_{I_m}}{\sigma^2} y_i - \frac{\omega}{2\sigma^2}\right) P(g_i|g_{im}, g_{if})}{\sum_{g_i^*, g_{im}^*, g_{if}^* \in G_\xi} \exp\left(\frac{\beta_{g_i^*}}{\sigma^2} y_i + \frac{\beta_{g_{im}^*}}{\sigma^2} y_i + \frac{\beta_{I_m}^*}{\sigma^2} y_i - \frac{\omega^*}{2\sigma^2}\right) P(g_i^*|g_{im}^*, g_{if}^*)} \end{aligned}$$

where

$$\begin{aligned} \omega^* &= \beta_{g_i^*}(2\alpha_M + \beta_{g_i^*}) + \beta_{g_{im}^*}(2\alpha_M + \beta_{g_{im}^*}) \\ &\quad + \beta_{I_m}(2\alpha_M + \beta_{I_m}^*) + 2\beta_{g_i^*}(\beta_{g_{im}^*} + \beta_{I_m}^*) + 2\beta_{g_{im}^*}\beta_{I_m}^* \end{aligned}$$