Comparison of Different Haplotype-Based Association Methods for Gene-Environment ($G \times E$) Interactions in Case-Control Studies when Haplotype-Phase Is Ambiguous

Rebecca Hein Lars Beckmann Jenny Chang-Claude

Department of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany

Key Words

Retrospective and prospective likelihood • Case-control studies • Indirect association • Expectation maximization algorithm • Hardy Weinberg equilibrium • Logistic regression • Linkage disequilibrium

Abstract

Objective: We compared four haplotype-based approaches for the analysis of gene-environment interactions when haplotype-phase is ambiguous. The methods employ different versions of the expectation maximization algorithm and differ in the choice of the reference group and in the way the risk of disease is modeled (retrospective versus prospective). Furthermore, the methods are based on distinct assumptions (such as Hardy Weinberg equilibrium). The haplotype-based methods were also compared to single-marker logistic regression (LR). **Methods:** We simulated case-control scenarios where the risk variant was directly genotyped (direct scenario) or in linkage disequilibrium with the genotyped markers (indirect scenario). **Results:** The retrospective methods tended to be more powerful for detecting interactions than the prospective methods. In the indirect scenarios, the power of all methods was decreased. However, the power of the retrospectives methods was high in some scenarios and the interactions may only be detectable when

using these approaches. Furthermore, we observed that the precision of one prospective method was clearly lower than the precision of the retrospective methods. **Conclusion:** For the analysis of gene-environment ($G \times E$) interactions in case-control data, the investigated retrospective methods can be an attractive alternative to haplotype-based methods which do not account for the retrospective sampling design.

Copyright © 2009 S. Karger AG, Basel

Introduction

Complex diseases such as cancer, obesity, and cardiovascular disease, comprise the main health burden in industrialized countries and thus are of great public health interest. These diseases are caused by the influence and interplay of genetic and environmental factors [1]. Thus, complex diseases are frequently being analyzed taking into account interactions between genetic and environmental factors, i.e., the modification of the effect of a genetic risk factor by one or more non-genetic risk factors or vice versa. The investigation of $G \times E$ interactions may provide greater etiologic insight into complex diseases. It allows the identification of differential effects, e.g. genetic effects that differ in exposed and unexposed individu-

als and is rewarding for the detailed characterization of identified disease genes (i.e., at advanced stages of genetic research).

In genetic association studies, haplotypes are frequently used to investigate the association between potential genetic risk factors and the disease of interest. Haplotypes are the arrangement of alleles on two or more markers in cis-position that are likely to be transmitted as a unit. Haplotypes may either comprise one or more disease causing variants, are in linkage disequilibrium (LD) with those variants, or are themselves functional. In contrast to single markers, haplotypes capture information on pair-wise as well as on multi-marker LD between true disease variants and genotyped markers. Thus, especially in indirect association studies, where the true risk variant or haplotype is not directly genotyped, haplotype-based approaches may be more powerful than single-marker methods [2].

The direct molecular determination of haplotypes in the laboratory is expensive and is hence done rarely. Thus, to estimate haplotype frequencies in a given study sample, haplotypes usually need to be derived from unphased multi-marker genotypes. A haplotype and its corresponding frequency can be uniquely determined from a multi-marker genotype only if the multi-marker genotype comprises at most one heterozygous marker. To estimate the phase of ambiguous haplotypes, different statistical algorithms have been developed. Among these algorithms [see e.g. 3–7], the expectation maximization (EM) algorithm [4, 5] is probably most frequently used.

In recent years, several statistical regression approaches have been developed for the haplotype-based analysis of $G \times E$ interactions when haplotype-phase is unknown. We focus here on four of these approaches for which software implementations are freely available: the two prospective approaches of Lake et al. [8] and Trégouët et al. [9, 10], and the retrospective methods of Lin et al. [11, 12] and Chen et al. [13]. Being publicly available, these programs are likely to be more commonly employed.

Retrospective approaches model the probability of genetic and environmental covariates conditional on disease status. By contrast, prospective methods are based on the joint likelihood of disease and genetic and environmental covariates. In retrospective case-control studies, individuals are retrospectively sampled with respect to their disease status, independently from the exposure to the genetic and environmental factors of interest. Unlike prospective methods, retrospective methods account for this retrospective sampling design of case-control studies. Thus, for the analysis of retrospective

case-control data, retrospective methods may be superior to prospective methods [13]. However, Prentice and Pyke [14] showed that retrospective and prospective likelihoods evaluated in retrospective data are equivalent if the covariate distribution, i.e., the distribution of the genetic and environmental factors, is left unrestricted as it is in traditional logistic regression (LR) analysis. Furthermore, Rabinowitz [15] and Breslow et al. [16] demonstrated that the prospective analysis of case-control data is efficient.

To estimate haplotype frequencies and regression parameters in haplotype-based approaches, assumptions about the distribution of covariates are required [17]. However, if the covariate distributions of the genetic and environmental variables are restricted by different assumptions the equivalence of prospective and retrospective likelihoods does not hold. Hardy-Weinberg equilibrium (HWE) among haplotype pairs in the pooled sample of cases and controls or in controls only is a frequently used assumption. Additionally, if environmental factors are incorporated into the statistical model, assumptions about the relationship between haplotypes and environmental factors may be required, such as independence between haplotypes and environmental exposures in the general population. If these assumptions are met, retrospective methods can lead to major efficiency gains in the analysis of case-control studies [13]. However, departure from these assumptions may cause serious bias when applying retrospective methods [18–20]. Satten and Epstein [19] compared prospective and retrospective methods for the analysis of haplotype main effects in retrospective case-control data. Their results demonstrate that prospective methods are less sensitive to violations of the HWE assumption than are retrospective methods.

Until now, comparisons of haplotype-based methods for the analysis of $G \times E$ interactions have not been conducted in the setting of either direct or indirect association studies. Using simulated case-control data, the characteristics of the four methods for the haplotype-based analysis of $G \times E$ interactions mentioned above and single-marker LR were compared. Direct and indirect scenarios as well as HWE and Hardy Weinberg disequilibrium (HWD) among haplotype pairs was simulated. Our results provide information for scientists in the process of deciding which methods to use for the analysis of their data, such that (1) the number of false positive results abides by the nominal significance level; (2) the probability to detect true positives is high, and (3) the precision of the effect estimates is high.

Methods

Prospective Haplotype-Trait Association Tests

The prospective haplotype-based methods are based on the joint probability of disease and genetic and environmental covariates: $P(Y, G, E)$, where Y represents the phenotype and G and E represent the covariates (G : unphased multi-marker genotype, E : coding for the environmental exposure). This joint probability can be decomposed into the prospective penetrance function (i.e., the conditional probability of disease given the covariates), $P(Y | G, E)$, and the probability of the covariates, $P(G, E)$.

For the haplotypes H , which need to be derived from the unphased multi-marker genotypes G , both prospective approaches assume that conditional on the individuals' haplotypes, the corresponding phenotypes Y are independent of their multi-marker genotypes, i.e., $P(Y, G | H) = P(Y | H) * P(G | H)$.

The prospective methods illustrated below estimate haplotype frequencies in the whole sample of cases and controls. Correspondingly, HWE among haplotype pairs is assumed in the whole sample.

The Prospective Approach of Lake et al. [8]

Schaid et al. [21] developed a prospective score test for the analysis of haplotype main effects in association studies. The approach is based on a generalized linear model. Lake et al. extended this method for the analysis of $G \times E$ interactions.

A weighted EM algorithm is employed to simultaneously estimate haplotype frequencies and regression parameters. In the maximization step (M-step) of the EM algorithm, the conditional expectation of the complete data log-likelihood is maximized to obtain maximum likelihood estimates (MLE) of the haplotype frequencies and the regression parameters. This conditional expectation depends on the joint distribution of the haplotype pairs and the marginal distribution of the haplotypes (i.e., the 'weights') conditional on the observed disease status, rather than on the true haplotype pairs derived from phase known multi-marker genotypes. In the expectation step (E-step), the conditional expectation of the complete data log-likelihood is calculated with respect to the observed data and the current estimates of haplotype frequencies and regression parameters.

All haplotypes and $G \times E$ interaction terms are included in the regression model with the exception of one reference haplotype – which usually is the most frequent one – and the corresponding interaction term. Haplotypes with frequencies below a fixed threshold, e.g. ≤ 0.05 , may be tested together in one group. The prospective haplotype-trait association test is implemented in the R-package *haplo.stats*, function: *haplo.glm* [8, 22, 23].

The Prospective Approach of Trégouet et al. [9, 10]

The prospective regression-based method of Trégouet et al. employs a stochastic version of the EM algorithm, the SEM algorithm. In each iteration, the distribution of haplotype pairs conditional on the observed data and the estimates of haplotype frequencies and regression parameters obtained from the previous iteration is simulated. In the stochastic E-step, each ambiguous haplotype pair is replaced by a haplotype pair drawn from the simulated distribution. In the M-step, these haplotype pairs are treated as if they were true. Updates of haplotype frequencies are derived by counting the number of haplotypes in the pseudo-

completed case-control sample. Regression parameters are then updated by standard maximum likelihood estimation.

The sequence of parameter estimates does not converge pointwise. Rather, it constitutes a Markov chain that converges to a stationary distribution. A burn-in period is required to obtain stationarity. The final parameter estimates are obtained by calculating the average of the parameter estimates obtained in all iterations which have been carried out after the burn-in period. Diebolt and Ip have shown that these estimates are asymptotically equivalent to the corresponding MLEs [24, 25]. Trégouet et al. state that due to its stochastic aspect, the SEM algorithm – in contrast to the traditional EM algorithm – is expected to be insensitive to local maxima or saddle points.

The reference group comprises only the most frequent haplotype and the corresponding $G \times E$ interaction term. The entirety of the remaining haplotypes – along with the corresponding $G \times E$ interaction terms – is included in the regression model. Command line and Java versions of the corresponding software *THE-SIAS* are freely available (see <http://ecgene.net/genecanvas/news.php>, [26]).

Retrospective Haplotype-Trait Association Tests

Epstein and Satten [17] developed a regression-based approach for the analysis of haplotype main effects when haplotype-phase is ambiguous, based on the retrospective likelihood function $P(G, E | Y)$. Both retrospective methods described below are extensions of this approach.

The Retrospective Approach of Lin et al. [11, 12]

Lin et al. introduced a retrospective haplotype-trait association test for $G \times E$ interactions that relies on the following presumptions: The environmental factors are assumed to be independent of the genotypes and haplotypes. In the likelihood function for the LR model a rare disease is presumed. Haplotype frequency estimates are obtained using either the pooled sample of cases and controls or the controls only. Correspondingly, the haplotype pair distribution of the respective sample is assumed to follow HWE. However, if the whole sample or the controls are not in HWE, this may be specified. The method accounts for HWD by incorporating the fixation index into the haplotype frequency estimation.

The dimension of the covariates may be high due to a large number of environmental factors and/or continuous environmental factors. Thus, maximization of retrospective likelihoods may be numerically challenging or even unfeasible [18]. To circumvent this potential problem, the profile likelihood approach is used: First the log-likelihood is maximized with respect to the probabilities of the environmental factors for fixed values of the haplotype frequencies and regression parameters; in a second step, the EM algorithm is applied using the profile likelihood derived in the first step. The E-step involves the evaluation of the expected log-likelihood using current parameter estimates. In the M-step, a one-step Newton-Raphson iteration is used to update the haplotype frequencies and the regression parameters simultaneously. Haplotype frequencies are estimated accounting for case-control status and all other covariates.

For the test of a specific haplotype or a $G \times E$ interaction effect, the reference group consists of all remaining haplotypes and the corresponding interaction terms. Furthermore, a global test for any haplotype main or $G \times E$ interaction effects can be per-

Table 1. Haplotype distribution

	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	Haplotype frequency
Haplotype 1 (h_1)	1	2	2	1	2	1	2	0.05
Haplotype 2 (h_2)	2	1	1	1	1	1	1	0.0625
Haplotype 3 (h_3)	1	2	1	1	2	1	2	0.0625
Haplotype 4 (h_4)	1	1	1	1	2	2	1	0.075
Haplotype 5 (h_5)	1	1	1	2	1	1	1	0.125
Haplotype 6 (h_6)	1	1	2	1	1	1	1	0.125
Haplotype 7 (h_7)	1	2	1	1	1	1	2	0.05
Haplotype 8 (h_8)	2	1	1	1	1	2	1	0.15
Haplotype 9 (h_9)	1	1	1	1	1	1	1	0.3

formed. Here, the reference is the most common haplotype and the corresponding interaction term. Accordingly, the regression model comprises all terms which do not belong to the reference group. The method was implemented in the software *HAPSTAT*, which can be downloaded from <http://www.bios.unc.edu/lin/hapstat/>.

The Retrospective Approach of Chen et al. [13]

Similar to Lin et al., Chen et al. introduced a retrospective approach for the haplotype-based analysis of G×E interactions. The method allows for dependence between genetic and environmental factors. However, the corresponding software implementation also accommodates the gene-environment independence assumption. Information on the disease prevalence or alternatively the rare disease approximation is required. Haplotypes are estimated in controls only and correspondingly the distribution of the haplotype pairs in controls is assumed to follow HWE. The haplotype pair distribution is generated using a polytomous LR model accounting for the environmental factors.

Similarly to the method of Lin et al., a profile likelihood approach is employed to profile out the probabilities of the environmental factors. The estimation procedure iterates between two steps until convergence. First, weights (i.e., probabilities for haplotype pairs conditional on disease status, genotype, and covariates) are calculated based on initial specifications of haplotype frequencies and the regression parameters. The second step solves the estimating equation based on the profile likelihood using the weights to update the haplotype frequencies and the regression parameters. The updated parameters are then used to recalculate the weights.

For each test, the reference group comprises one (common) haplotype, e.g. the most frequent one along with the corresponding G×E interaction term. Moreover, it is possible to add rare haplotypes to the reference group. All terms that are not included in the reference group can be incorporated into the regression model. This method was implemented in the SAS (SAS Institute Inc., Cary, N.C., USA) macro *HapReg* and is available for download (http://www.stat.sinica.edu.tw/yhchen/macro_hapreg.sas).

Simulation Scenarios

Individuals were simulated by drawing haplotype pairs and exposure status (prevalence of the environmental exposure: 0.1) according to a given distribution (table 1). The disease was modeled to be rare (baseline risk of disease: 0.017). Disease status was assigned using Bernoulli random variables with individual disease probabilities defined according to a LR model, conditional on the genotype at the risk variant and the environmental exposure:

$$\text{Logit}(Y|G,E) = \alpha + \beta_G * G + \beta_E * E + \beta_{GE} * G * E,$$

where α is the intercept ($\exp(\alpha)/(1 + \exp(\alpha))$: baseline risk of disease). β_G , β_E , and β_{GE} are the regression coefficients that correspond to the log-odds ratios of the genetic and environmental main effects and the magnitude of the G×E interaction, respectively. G is the code for the genotype at the risk variant, which depends on the mode of inheritance. For example, for a variant with risk allele A and non-risk allele a and that follows a multiplicative mode of inheritance (i.e., the risk in homozygous carriers of the risk allele is the square of the risk in heterozygous carriers), G is equal to 0 for genotype aa, 1 for genotype aA, and 2 for genotype AA. E – i.e., the coding for the environmental exposure – is equal to 1 if an individual is exposed and 0 otherwise.

To simulate data under the null hypothesis of no main and G×E interaction effects, the coefficients β_G , β_E , and β_{GE} in the LR model were fixed to 0. In the scenarios 1⁰, 2^{0,het}, and 2^{0,hom} (table 2), which were simulated under the null hypothesis, all haplotypes were assumed to be identically distributed among cases and controls.

Under the alternative hypothesis, i.e., in presence of main as well as G×E interaction effects, the risk variant

Table 2. Simulation scenarios

Scenario	Risk variant	Hypothesis	Mode of inheritance	Haplotype pair distribution
1 ⁰	–	null	–	HWE ^b
2 ^{0, het}	–	null	–	HWD ^c (excess heterozygosity)
2 ^{0, hom}	–	null	–	HWD ^c (excess homozygosity)
1 ^{dir}	directly genotyped	alternative	multiplicative	HWE ^b
1 ^{indir}	in LD ^a with genotyped markers	alternative	multiplicative	HWE ^b
2 ^{dir, het}	directly genotyped	alternative	multiplicative	HWD ^c (excess heterozygosity)
2 ^{indir, het}	in LD ^a with genotyped markers	alternative	multiplicative	HWD ^c (excess heterozygosity)
2 ^{dir, hom}	directly genotyped	alternative	multiplicative	HWD ^c (excess homozygosity)
2 ^{indir, hom}	in LD ^a with genotyped markers	alternative	multiplicative	HWD ^c (excess homozygosity)

⁰ = scenarios simulated under the null hypothesis; ^{dir} = scenarios simulated under the alternative hypothesis and the direct association approach; ^{indir} = scenarios simulated under the alternative hypothesis and the indirect association approach; ^{het} = heterozygosity; ^{hom} = homozygosity.

^a Linkage disequilibrium; ^b Hardy Weinberg equilibrium; ^c Hardy Weinberg disequilibrium.

was defined to be SNP6 in table 1 (frequency of the risk allele: 0.225). The genetic and environmental main effects were chosen to be moderate ($\exp(\beta_G) = \exp(\beta_E) = 1.75$), whereas the G×E interaction effect was chosen to be strong (magnitude of the interaction: $\exp(\beta_{GE}) = 3$). The risk haplotypes were identified as those that harbor the risk allele, i.e., haplotypes h_4 and h_8 (table 1). Both direct and indirect association approaches were considered. In the direct scenarios 1^{dir}, 2^{dir, het}, and 2^{dir, hom} (table 2), haplotypes comprised seven markers including the disease variant and were uniquely identifiable. For indirect association analysis, the risk variant was removed from the data in scenarios 1^{indir}, 2^{indir, het}, and 2^{indir, hom} (table 2). Thus, in the resulting haplotype population, risk haplotype h_4 was uniquely distinguishable, whereas risk haplotype h_8 was identical to the non-risk haplotype h_2 . Under the alternative hypothesis, all scenarios were generated under a multiplicative mode of inheritance. Therefore, in scenarios 1^{dir} and 1^{indir}, the haplotypes were in HWE in controls as well as in cases.

HWE may be distorted by genotyping errors or by population genetic forces such as inbreeding, mutation, selection, and gene flow. Whereas inbreeding may lead to excess of homozygous haplotype pairs, selection may induce excess of heterozygote haplotype pairs. To investigate the performance of the methods under HWD in the whole sample, a weighted Dirichlet distribution was used to simulate frequencies of haplotype pairs that are skewed away from HWE in the manner described by Niu et al. [7]. For the nine haplotypes in table 1, the 9×9 matrix $C = (c_{ij})_{9 \times 9}$ was modeled such that it comprised the probabilities

c_{ij} of picking haplotype pair (h_i, h_j) (i.e., $c_{ij} \geq 0$, and $\sum c_{ij} = 1$). C followed the Dirichlet distribution *Dirichlet*(100D). $D = (d_{ij})_{9 \times 9}$ was a probability matrix and the elements of D were given by $d_{ij} = w_{ij} * p_i * p_j$, p_i denoting the frequency of haplotype h_i . When all elements of the matrix $W = (w_{ij})_{9 \times 9}$ were equal to 1, the matrix W gave equal weights to all haplotype pairs and thus the haplotype pairs in the simulated sample were in HWE. To increase the number of heterozygous haplotype pairs, more weight was given to dissimilar haplotype pairs. The diagonal elements of the matrix W were then set equal to 1, whereas all other elements were fixed to 3 to simulate excess heterozygosity (scenarios 2^{0, het}, 2^{dir, het}, and 2^{indir, het}). Correspondingly, excess of homozygous haplotype pairs was simulated by fixing the diagonal elements of matrix W to 3 and the remaining elements of matrix W to 1 (scenarios 2^{0, hom}, 2^{dir, hom}, and 2^{indir, hom}). These HWD models have also been used by Niu et al. [7] and Lake et al. [8], but differ from the models used in the method of Lin et al. when accounting for HWD.

For each scenario described above (see also table 2), 1,000 replications – each comprising 400 case-control pairs – were produced under the null and the alternative hypotheses.

Data Analysis

The simulation scenarios outlined above were analyzed with respect to family wise error rate (FWER), power, bias, variance, coverage probability (CP) for 95%

confidence intervals, and asymptotic relative efficiency (ARE) using single-marker LR as well as the haplotype-based methods described above. The LR models consisted of one single variant and the corresponding G×E interaction term. For each of the variants presented in table 1 (i.e., the 7 variants SNP1, ..., SNP7 in the direct scenarios and the 6 variants SNP1, ..., SNP5, SNP7 in the indirect scenarios), two tests, i.e., one test for the main effect and one test for the interaction, were performed. Accordingly, in the haplotype-based analyses, all common haplotypes (risk and non-risk haplotypes) with frequency ≥ 0.05 and the respective interaction terms were separately tested for association to the disease. The regression model of the method of Lin et al. comprised one haplotype and the corresponding G×E interaction term at a time. The corresponding reference group consisted of all terms except for the tested haplotype and the G×E interaction term. For technical reasons and/or convergence problems, it was not possible to use the same model for the other methods. Therefore, the methods of Lake et al., Trégouet et al., and Chen et al. were applied such that the regression model contained all haplotypes and G×E interaction terms except for haplotype h_9 (table 1), i.e., the most frequent haplotype and, if applicable, for the group of rare haplotypes, which formed the reference group. All data analyses were performed assuming a multiplicative mode of inheritance. When haplotype pairs were simulated to deviate from HWE, the method of Lin et al. was applied twice: under the assumption of HWE and of HWD. The other haplotype-based approaches do not facilitate the assumption of HWD and therefore were consistently applied under the assumption of HWE. The method of Lin et al. was applied such that haplotypes were estimated in the pooled sample rather than in controls only. In the application of the methods of Lin et al. and Chen et al., gene-environment independence and a rare disease were assumed, which are valid assumptions in the simulated data.

In our paper, we were interested in the power of the methods to detect the specific risk haplotype(s) or variant(s) in the genetic region. Therefore, for the haplotype-based methods, power was calculated as the percentage of replicates in which the main or interaction effect of the respective risk haplotype had a p value $\leq 0.05/(2 \times \text{nhap})$. Here, nhap is the number of haplotypes with frequency ≥ 0.05 in the respective replication of each particular scenario. For LR analysis, power was calculated as the percentage of replicates in which the p value of the main or interaction effect of the risk variant SNP6, in the direct scenarios, or the surrogate variant SNP1, in the indirect scenarios, was $\leq 0.05/$

(2×nSNPs). The extent of LD between risk variant SNP6 and surrogate variant SNP1 is demonstrated in table 1 and 2 in the online supplementary material (www.karger.com/doi/10.1159/000228923). In the direct and indirect scenarios, nSNPs (i.e., the number of tested variants) was 7 and 6, respectively. The FWER was derived as the percentage of replicates in which any main or interaction effect had a p value $\leq 0.05/(2 \times \text{nhap})$ for the haplotype-based methods and $\leq 0.05/(2 \times \text{nSNPs})$ for LR. For each main and G×E interaction effect, the mean bias/variance/CP was estimated as the mean of the biases/variances/CPs of the respective effect estimates over all replications. For example, the mean variance was calculated as

$$\bar{\sigma}^2 = \frac{1}{nrep} * \sum_{i=1}^{nrep} \hat{\sigma}_i^2,$$

where $\hat{\sigma}_i^2$ is the variance estimated in replication i and $nrep$ is the number of replications. Additionally, in the online supplementary tables 6, 7, and 13, we present the sample variance of the parameter estimates of all replications, i.e.,

$$\hat{\sigma}_\theta^2 = \frac{1}{nrep - 1} * \sum_{i=1}^{nrep} (\hat{\theta}_i - \bar{\theta})^2,$$

where $\hat{\theta}_i$ is the parameter estimated in replication i and $\bar{\theta}$ is the mean of the parameter estimates of the respective effect over all replications. The ARE was calculated as the ratio of the mean variance of a haplotype or G×E interaction effect of a prospective haplotype-based method relative to the corresponding mean variance of a retrospective haplotype-based method.

Additionally, we evaluated the performance of the four haplotype-based methods with respect to haplotype frequency estimation. Over all replications, we calculated the mean of (1) the sum of absolute differences between estimated and true frequencies (SAD) and (2) the sum of squared errors between estimated and true frequencies (SSE):

$$SAD = \sum_{n=1}^{nhap} |\hat{f}_n - f_n|$$

and

$$SSE = \sum_{n=1}^{nhap} (\hat{f}_n - f_n)^2,$$

where $\hat{f}_n$ and f_n are the estimated and true frequency corresponding to the n -th haplotype [27]. Differences in SAD and SSE between the prospective versus retrospective methods were assessed using the two-sided paired t test.

Table 3. Haplotype certainty measured in terms of sum of (1) absolute and (2) squared differences between estimated and true haplotype frequencies for the methods of Lake et al., Trégouet et al., Lin et al., and Chen et al. for the scenarios simulated under the null hypothesis: 1^0 (Hardy Weinberg equilibrium) and $2^{0, \text{het}}$ and $2^{0, \text{hom}}$ (Hardy Weinberg disequilibrium)

Scenario		1^0	$2^{0, \text{het}}$	$2^{0, \text{hom}}$
f_h^a	h_4	0.10	0.10	0.10
	h_8	0.13	0.11	0.14
Lake et al. (prospective)	SAD ^d (CI ^f)	0.0003 (0.0001, 0.0006)	$<10^{-4}$	0.0003 ($<10^{-4}$, 0.0006)
	SSE ^e (CI ^f)	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
Trégouet et al. (prospective)	SAD ^d (CI ^f)	0.0003 ($<10^{-4}$, 0.0005)	$<10^{-4}$	0.0004 ($<10^{-4}$, 0.0008)
	SSE ^e (CI ^f)	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
Lin et al. (retrospective) ^b	SAD ^d (CI ^f)	0.0003 ($<10^{-4}$, 0.0005)	$<10^{-4}$	0.0003 ($<10^{-4}$, 0.0006)
	SSE ^e (CI ^f)	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
Lin et al. (retrospective) ^c	SAD ^d (CI ^f)	–	$<10^{-4}$	0.0005 (0.0002, 0.0008)
	SSE ^e (CI ^f)	–	$<10^{-4}$	$<10^{-4}$
Chen et al. (retrospective)	SAD ^d (CI ^f)	0.0009 (0.0005, 0.0014)	$<10^{-4}$	0.0005 (0.0001, 0.0009)
	SSE ^e (CI ^f)	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$

^a Haplotype frequency; ^b method of Lin et al. when not accounting for Hardy Weinberg disequilibrium; ^c method of Lin et al. when accounting for Hardy Weinberg disequilibrium; ^d sum of absolute differences between estimated and true haplotype frequencies rounded to four digits; ^e sum of squared errors between estimated and true haplotype frequencies rounded to four digits, and ^f confidence interval.

Results

Convergence

In the presented analyses, convergence was an issue due to the rare haplotypes estimated from the simulated data. Whereas the prospective method of Lake et al. and the retrospective method of Lin et al. failed to converge in less than 0.001% of the analyses, the prospective method of Trégouet et al. and the retrospective method of Chen et al. did not converge in 15 and 14% of the analyses, respectively. Thus, FWER, power, bias, variance, CP and ARE calculations took into account only those analyses that did not have any convergence problems.

Haplotype Estimation

In table 3 and in the online supplementary tables 3 and 10, haplotype certainty is quantified in terms of SAD and SSE, which measure the overall discrepancy between true and estimated haplotype frequencies. In these tables, the haplotype frequencies presented for haplotypes h_4 and h_8 are the haplotype frequencies for the first replication of the respective scenarios. We observed that SAD and SSE were comparably small (i.e., haplotype certainty was high) for the methods of Lake et al., Trégouet et al. and Lin et al., and did not exceed 0.01, except for the scenario 1^{indir} with maximum SAD of 0.013. For the scenarios 1^{dir}

and 1^{indir} , the method of Chen et al. showed a relatively high value for SAD (0.0982 and 0.0865), while the performance of the method was comparable to the others in the remaining scenarios. When conducting two-sided paired t tests for the differences between SAD/SSE values of prospective versus retrospective methods, significant differences (i.e., p values ≤ 0.05) were observed in some scenarios (online suppl. tables 4, 5, and 11).

Scenarios Simulated Under the Null Hypothesis

In table 4 (see also online suppl. fig. 1), the FWER for main and interaction effects of single-marker LR and the four haplotype-based approaches for the analysis of $G \times E$ interactions in case-control studies is presented for scenario 1^0 , simulated under HWE, and scenarios $2^{0, \text{het}}$ and $2^{0, \text{hom}}$, simulated under the two HWD scenarios described above. Additionally, table 4 exemplarily presents the bias, mean variance, and CP for the risk variant SNP6, in LR analysis, and for haplotype h_4 , in haplotype-based analysis, for these scenarios. For the haplotype-based approaches, the corresponding ARE is shown in table 5.

All methods (single-marker and haplotype-based) generally tended to be conservative in scenarios 1^0 , $2^{0, \text{het}}$, and $2^{0, \text{hom}}$ (table 4). However, under excess homozygosity (scenario $2^{0, \text{hom}}$), the retrospective method of Lin et al. reflected the nominal significance level when applied

Table 4. Family wise error rate (FWER) as well as bias, mean variance and coverage probability (CP) for SNP6 (logistic regression) and haplotype h_4 (see table 1; haplotype-based analysis: the methods of Lake et al., Trégouet et al., Lin et al., and Chen et al.) for simulation scenarios 1^0 (Hardy Weinberg equilibrium) and $2^{0, \text{het}}$ and $2^{0, \text{hom}}$ (Hardy Weinberg disequilibrium)

Scenario			1^0	$2^{0, \text{het}}$	$2^{0, \text{hom}}$
Logistic regression	FWER ^a	main	0.02	0.01	0.02
		int.	0.02	0.01	0.00
	bias	main	-0.74	-0.75	-0.74
		int.	-1.92	-1.89	-1.92
	variance	main	1.02	1.02	1.01
		int.	1.20	1.21	1.17
	CP ^b	main	0.00	0.00	0.00
		int.	0.01	0.00	0.00
Lake et al. (prospective)	FWER ^a	main	0.02	0.02	0.03
		int.	0.01	0.01	0.00
	bias	main	0.04	0.03	0.02
		int.	-1.57	-1.45	-1.66
	variance	main	1.05	1.06	1.04
		int.	1.92	2.01	1.74
	CP ^b	main	0.00	0.00	0.00
		int.	0.02	0.02	0.02
Trégouet et al. (prospective)	FWER ^a	main	0.02	0.03	0.02
		int.	0.01	0.01	0.01
	bias	main	0.03	0.03	0.02
		int.	0.94	0.94	0.95
	variance	main	0.94	0.94	0.95
		int.	1.70	1.83	1.54
	CP ^b	main	0.00	0.00	0.00
		int.	0.02	0.02	0.02
Lin et al. (retrospective)	FWER ^{a, c}	main	0.03	0.02	0.05
		int.	0.03	0.02	0.05
	bias ^c	main	0.02	0.01	0.02
		int.	1.02	1.03	1.03
	variance ^c	main	1.02	1.03	1.03
		int.	1.18	1.18	1.20
	CP ^{b, c}	main	0.10	0.07	0.10
		int.	0.69	0.74	0.69
	FWER ^{a, d}	main	-	0.02	0.07
		int.	-	0.02	0.03
	bias ^d	main	-	0.01	0.01
		int.	-	1.03	1.02
	variance ^d	main	-	1.03	1.02
		int.	-	1.18	1.17
	CP ^{b, d}	main	-	0.07	0.05
		int.	-	0.74	0.67
Chen et al. (retrospective)	FWER ^a	main	0.02	0.01	0.10
		int.	0.01	0.00	0.06
	bias	main	-0.72	-0.73	-0.72
		int.	-1.87	-1.89	-1.83
	variance	main	1.05	1.05	1.04
		int.	1.40	1.42	1.33
	CP ^b	main	0.00	0.00	0.00
		int.	0.01	0.00	0.01

^a Family wise error rate; ^b coverage probability for the 95% confidence intervals; ^c method of Lin et al. when not accounting for Hardy Weinberg disequilibrium, and ^d method of Lin et al. when accounting for Hardy Weinberg disequilibrium.

Table 5. Asymptotic relative efficiency of the prospective methods of Lake et al. and Trégouet et al. versus the retrospective methods of Lin et al. and Chen et al. for haplotype h_4 in scenarios 1^0 (Hardy Weinberg equilibrium) and $2^{0, \text{het}}$ and $2^{0, \text{hom}}$ (Hardy Weinberg disequilibrium) simulated under the null hypothesis

Scenario		1^0	$2^{0, \text{het}}$	$2^{0, \text{hom}}$
Lake (prospective) versus Lin (retrospective)	Main	1.03	1.03	1.01
	Int.	1.63	1.71	1.45
Lake (prospective) versus Chen (retrospective)	Main	1.00	1.00	1.00
	Int.	1.37	1.41	1.31
Trégouet (prospective) versus Lin (retrospective)	Main	0.92	0.91	0.92
	Int.	1.45	1.55	1.29
Trégouet (prospective) versus Chen (retrospective)	Main	0.89	0.89	0.91
	Int.	1.22	1.28	1.16

under the assumption of HWE. Similarly, for the interaction effect, the retrospective method of Chen et al. abided by the nominal value in scenario $2^{0, \text{hom}}$, but was too liberal for haplotype main effects.

The bias of the estimated genetic (SNP or haplotype) main effects was negligible for all approaches except for the retrospective method of Chen et al. With this method, the haplotype main effect was slightly underestimated, i. e., a protective effect (<1) was obtained. Corresponding estimates of the mean variance were about 1 for all methods (table 4) and the sample variance of the parameter estimates of all replications was small (online suppl. table 6). The interaction effect was underestimated by the LR approach as well as the prospective method of Lake et al. and the retrospective approach of Chen et al. (bias in the range of -1.45 to -1.92). In contrast, the haplotype-based methods of Trégouet et al. and Lin et al. overestimated this effect (bias about 1). Compared to the variance estimates for the main effects, the variance estimates of the interaction effects were generally increased. While the variances were only slightly increased for LR and the retrospective method of Lin et al. and moderately increased for the retrospective method of Chen et al., they were considerably increased for the prospective methods of Lake et al. and Trégouet et al. (table 4, online suppl. table 6).

The CP was generally very low, i.e., between 0.00 and 0.10. However, for G×E interactions, the CP derived from the retrospective method of Lin et al. was in the range of 0.67 to 0.74 (table 4).

For the analysis of haplotype main effects, the precision of the prospective method of Lake et al. was comparable to the precision of both retrospective approaches, i.e., the ARE was about 1 (table 5). Furthermore, the pre-

cision of the prospective method of Trégouet et al. for haplotype main effects was slightly higher than that of the two retrospective methods (ARE values of about 0.90 in table 5). However, for the analysis of G×E interaction effects, the precision of the retrospective approaches was higher than the precision of the prospective methods (reflected by ARE values in the range of 1.16 to 1.71). This result is in line with the observation that the variances of the interaction effects estimated with the retrospective methods are smaller than the variances estimated with the prospective haplotype-based approaches. Furthermore, the ARE values reflect, that the precision of the method of Lin et al. was somewhat higher than the precision of the method of Chen et al.

In scenarios $2^{0, \text{het}}$ and $2^{0, \text{hom}}$, the method of Lin et al. was applied twice, i.e., under the assumption of HWE as well as under the assumption of HWD. For scenario $2^{0, \text{het}}$, simulated under excess heterozygosity, the fixation index was estimated to be unequal to zero in less than 1% of the replicates, i.e., more than 99% of the analyses were identical to those conducted under the assumption of HWE. By contrast, when excess homozygosity was simulated (scenario $2^{0, \text{hom}}$), the fixation index in all analyses was estimated to be unequal to zero and was thus incorporated into haplotype frequency estimation. However, the performance of the method was observed to be similar in both HWD scenarios.

Scenarios Simulated under the Alternative Hypothesis

Table 6 (see also online suppl. fig. 2) presents the power, bias, mean variance, and CP for the main and G×E interaction effects, when data were simulated under a multiplicative mode of inheritance. For LR, the presented values in table 6 correspond to the risk variant SNP6 (see table 1) in the direct scenarios (1^{dir} , $2^{\text{dir, het}}$, $2^{\text{dir, hom}}$) and to the surrogate variant SNP1 in the indirect scenarios (1^{indir} , $2^{\text{indir, het}}$, $2^{\text{indir, hom}}$). The results of the haplotype-based analyses (the prospective methods of Lake et al. and Trégouet et al. and the retrospective methods of Lin et al. and Chen et al.) are represented for the two risk haplotypes h_4 and h_8 . Whereas h_4 is uniquely distinguishable in both direct scenarios and indirect scenarios, h_8 is not distinguishable from haplotype h_2 (table 1) in the indirect scenarios.

In the direct scenario 1^{dir} , simulated under HWE, the power to detect the main effect of haplotype h_4 , with low haplotype frequency, was only low to moderate for all haplotype-based methods (table 6). By contrast, the power to identify the main effect of haplotype h_8 , with moderate haplotype frequency, was moderate to high. A very high power of 0.93 to detect the SNP main effect in sce-

nario 1^{dir} was obtained with LR. Among the haplotype-based methods, the highest power to detect the main effects of the two risk haplotypes h_4 and h_8 in scenario 1^{dir} was obtained with the prospective method of Trégouet et al. The power of the prospective methods to detect G×E interactions was lower compared to the power of these methods to identify haplotype main effects. By contrast, the power of the retrospective methods was on average higher to detect G×E interaction versus haplotype main effects. For the identification of G×E interactions in the different scenarios, the retrospective method of Lin et al. tended to be more powerful than both of the other haplotype-based approaches and the single-marker LR. It was merely less powerful than LR to detect the interaction of haplotype h_4 with the environmental exposure in scenario 1^{dir} . Furthermore, the power of the method of Chen et al. to identify interaction effects was inclined to be higher than the power of the prospective haplotype-based methods. In the indirect scenarios (1^{indir} , $2^{\text{indir, het}}$, $2^{\text{indir, hom}}$), the power of all methods to detect risk haplotype h_8 and the corresponding G×E interaction was generally lower than in the direct scenarios. However, the power to detect risk haplotype h_4 (which was uniquely identifiable in the indirect scenarios) as well as the related G×E interaction effect was similar or even higher in the indirect compared to the direct scenarios. When HWD was simulated (scenarios $2^{\text{dir, het}}$ and $2^{\text{indir, het}}$, $2^{\text{dir, hom}}$, and $2^{\text{indir, hom}}$), the power of the different methods was generally lower compared to the scenarios simulated under HWE. However, the power of the method of Lin et al. to detect the main as well as the G×E interaction effect of haplotype h_4 was higher in the scenarios simulated under HWD compared to scenarios 1^{dir} and 1^{indir} , which were simulated under HWE. In the scenarios simulated under HWD, the power of the retrospective method of Lin et al. was higher than the power of the other haplotype-based approaches.

For all methods, the bias of the main effect was smaller (in the range of -1.21 to 0.37) than the bias observed for the interaction effect (in the range of -1.84 to 6.77). On average, largest biases for the main and interaction effects were observed for the retrospective method of Chen et al. When HWD was simulated (scenarios $2^{\text{dir, het}}$, $2^{\text{indir, het}}$, $2^{\text{dir, hom}}$, and $2^{\text{indir, hom}}$), the bias of the main and interaction effects were increased on average. However, this does not apply to the bias of the G×E interaction estimated with the prospective method of Trégouet et al. and the retrospective method of Lin et al.

The mean variance estimates of the prospective method of Trégouet et al. varied strongly among the different scenarios and were very small (i.e., close or equal to zero)

Table 6. Power, bias, mean variance, and coverage probability for the variants SNP6 and SNP1 (logistic regression; direct and indirect scenarios) and the haplotypes h_4 and h_8 (see haplotype-based methods: the methods of Lake et al., Trégouet et al., Lin et al., and Chen et al.) for the scenarios simulated under a multiplicative mode of inheritance: 1^{dir} and 1^{indir} (Hardy Weinberg equilibrium) and $2^{\text{dir, het}}$, $2^{\text{indir, het}}$, $2^{\text{dir, hom}}$, and $2^{\text{indir, hom}}$ (Hardy Weinberg disequilibrium)

Scenario			1^{dir}	1^{dir}	1^{indir}	1^{indir}	$2^{\text{dir, het}}$	$2^{\text{dir, het}}$	$2^{\text{indir, het}}$	$2^{\text{indir, het}}$	$2^{\text{dir, hom}}$	$2^{\text{dir, hom}}$	$2^{\text{indir, hom}}$	$2^{\text{indir, hom}}$
SNP			SNP6		SNP1		SNP6		SNP1		SNP6		SNP1	
Logistic regression	power	main	0.93		0.41		0.87		0.09		0.96		0.09	
		int.	0.55		0.11		0.44		0.04		0.38		0.01	
	bias	main	0.03		-0.31		0.01		-0.51		0.03		-0.54	
		int.	0.27		-0.92		0.12		-1.38		-0.20		-1.56	
	variance	main	1.02		1.02		1.02		1.02		1.02		1.02	
		int.	1.15		1.14		1.16		1.14		1.14		1.12	
	CP ^a	main	0.06		0.02		0.05		0.00		0.04		0.00	
		int.	0.06		0.03		0.06		0.01		0.08		0.00	
Haplotype			h_4	h_8	h_4	h_8	h_4	h_8	h_4	h_8	h_4	h_8	h_4	h_8
Lake et al. (prospective)	power	main	0.39	0.66	0.22	0.23	0.02	0.01	0.06	0.00	0.09	0.12	0.09	0.00
		int.	0.09	0.24	0.05	0.11	0.01	0.01	0.01	0.00	0.01	0.03	0.00	0.00
	bias	main	0.08	0.05	-0.21	-0.19	-0.62	-0.68	-0.42	-1.04	-0.38	-0.39	-0.39	-0.82
		int.	1.46	0.70	0.15	-0.08	-1.40	-1.39	-1.10	-1.81	-0.83	-0.43	-1.27	-0.80
	variance	main	1.05	1.03	1.05	1.03	1.06	1.06	1.05	1.04	1.05	1.05	1.04	1.03
		int.	1.59	1.31	1.59	1.24	1.58	1.56	1.45	1.40	1.51	1.55	1.37	1.28
	CP ^a	main	0.05	0.06	0.04	0.04	0.01	0.00	0.02	0.00	0.01	0.01	0.02	0.00
		int.	0.06	0.08	0.03	0.05	0.03	0.02	0.03	0.00	0.00	0.02	0.01	0.00
Trégouet et al. (prospective)	power	main	0.40	0.73	0.25	0.28	0.02	0.01	0.06	0.00	0.08	0.11	0.09	0.00
		int.	0.22	0.01	0.14	0.01	0.02	0.02	0.03	0.00	0.03	0.00	0.01	0.00
	bias	main	0.07	0.04	-0.18	-0.19	-0.65	-0.69	-0.42	-1.00	-0.63	-0.61	-0.47	-0.94
		int.	1.05	1.03	1.05	0.57	0.52	0.59	1.04	0.00	0.64	0.64	1.02	0.02
	variance	main	1.05	1.03	1.05	0.57	0.52	0.59	1.04	0.00	0.64	0.64	1.02	0.02
		int.	1.32	1.98	1.31	1.06	0.85	1.11	1.51	0.00	0.95	1.19	1.36	0.04
	CP ^a	main	0.09	0.06	0.03	0.02	0.00	0.00	0.02	0.00	0.01	0.01	0.02	0.00
		int.	0.07	0.02	0.02	0.02	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Lin et al. (retrospective)	power ^{a, b}	main	0.25	0.67	0.26	0.41	0.40	0.35	0.36	0.07	0.57	0.53	0.62	0.13
		int.	0.39	0.86	0.41	0.67	0.50	0.55	0.51	0.14	0.76	0.68	0.86	0.15
	bias ^b	main	-0.14	-0.07	-0.15	-0.3	-0.18	-0.2	-0.2	-0.53	0.04	0.05	0.04	-0.43
		int.	1.04	1.02	1.04	1.02	1.03	1.03	1.03	1.02	1.03	1.03	1.03	1.02
	variance ^b	main	1.04	1.02	1.04	1.02	1.03	1.03	1.03	1.02	1.03	1.03	1.03	1.02
		int.	1.05	1.03	1.05	1.03	1.04	1.04	1.04	1.03	1.04	1.04	1.04	1.03
	CP ^{ab}	main	0.93	0.94	0.91	0.67	0.88	0.87	0.88	0.18	0.94	0.93	0.92	0.27
		int.	0.96	0.94	0.96	0.95	0.96	0.96	0.95	0.69	0.85	0.89	0.79	0.64
	power ^c	main	-	-	-	-	0.40	0.35	0.36	0.07	0.47	0.38	0.56	0.11
		int.	-	-	-	-	0.50	0.55	0.51	0.14	0.68	0.57	0.79	0.11
	bias ^c	main	-	-	-	-	-0.18	-0.20	-0.20	-0.53	-0.14	-0.14	-0.18	-0.49
		int.	-	-	-	-	1.03	1.03	1.03	1.02	1.02	1.02	1.02	1.01
	variance ^c	main	-	-	-	-	1.03	1.03	1.03	1.02	1.02	1.02	1.02	1.01
		int.	-	-	-	-	1.04	1.04	1.04	1.03	1.03	1.03	1.03	1.02
	CP ^{ac}	main	-	-	-	-	0.88	0.87	0.88	0.18	0.76	0.70	0.83	0.11
		int.	-	-	-	-	0.96	0.96	0.95	0.69	0.95	0.97	0.95	0.51
Chen et al. (retrospective)	power	main	0.38	0.57	0.42	0.28	0.08	0.00	0.02	0.00	0.14	0.14	0.19	0.00
		int.	0.64	0.53	0.68	0.22	0.44	0.01	0.02	0.00	0.19	0.37	0.19	0.01
	bias	main	0.09	0.20	-0.46	0.37	-1.21	-0.69	-0.44	-1.11	-0.36	-1.00	-0.33	-0.99
		int.	-0.46	-0.87	-1.38	-1.11	6.77	-1.84	-1.72	3.07	-1.34	4.22	-1.51	2.45
	variance	main	1.05	1.05	1.04	1.08	1.14	1.06	1.05	1.09	1.05	1.13	1.04	1.11
		int.	1.08	1.11	1.07	1.12	1.18	1.08	1.07	1.14	1.07	1.16	1.05	1.12
	CP ^a	main	0.06	0.03	0.03	0.04	0.00	0.01	0.01	0.00	0.02	0.01	0.04	0.00
		int.	0.05	0.03	0.03	0.02	0.01	0.00	0.00	0.00	0.02	0.04	0.00	0.00

^a Coverage probability for the 95% confidence intervals; ^b method of Lin et al. when not accounting for Hardy Weinberg disequilibrium; ^c method of Lin et al. when accounting for Hardy Weinberg disequilibrium

Fig. 1. Family wise error rate (FWER) of single-marker logistic regression (solid lines) and the four haplotype-based methods (dots) with respect to the $G \times E$ interaction effects in scenario 1^0 (simulated under HWE) and scenarios $2^{0, \text{het}}$, $2^{0, \text{hom}}$ (simulated under HWD). A: prospective method Lake et al.; B: prospective method of Trégouet et al.; C: retrospective method of Lin et al. (not accounting for HWD); D: retrospective method of Lin et al. (accounting for HWD), and E: retrospective method of Chen et al.

in some scenarios (table 6; especially for haplotype h_8 in scenarios $2^{\text{indir, het}}$ and $2^{\text{indir, hom}}$). The mean variance estimates obtained by all other approaches for the genetic main effect were close to 1. Compared to the variances estimated for haplotype main effects, the variances for the $G \times E$ interactions were moderately increased when applying the prospective method of Lake et al. (i.e., mean variances in the range of 1.24 to 1.59) and slightly increased when using single-marker LR and the retrospective method of Chen et al. (i.e., mean variances in the range of 1.05 to 1.18). By contrast, the variance of the retrospective method of Lin et al. was comparable to the variance estimates in the analyses of haplotype main effects (i.e., vari-

ance estimates were approximately 1). A similar pattern was observed for the sample variance of the parameter estimates of all replications (online suppl. table 7).

For all methods except the retrospective method of Lin et al., the CP was very low, i.e., in the range of 0 to 0.08. However, in the application of the method of Lin et al., the estimated 95% confidence intervals cover the true main and interaction effect on average 71 and 87% of the time.

The precision of the prospective method of Lake et al. was similar to the precision of the two retrospective approaches for the investigation of haplotype main effects (see online suppl. table 8), i.e., the ARE comparing this

Fig. 2. Power of single-marker logistic regression (solid lines) and the four haplotype-based methods (dots) with respect to the G×E interaction effects for the scenarios simulated under a multiplicative mode of inheritance: 1^{dir} and 1^{indir} (HWE) and 2^{dir, het}, 2^{indir, het}, 2^{dir, hom}, and 2^{indir, hom} (HWD). A: prospective method Lake et al.; B: prospective method of Trégouet et al.; C: retrospective method of Lin et al. (not accounting for HWD); D: retrospective method of Lin et al. (accounting for HWD), and E: retrospective method of Chen et al.

prospective approach with the two retrospective approaches was approximately 1. The precision of the prospective method of Trégouet et al. for tests of haplotype main effects tended to be higher than the precision of the retrospective approaches (average ARE: 0.66). For the analysis of G×E interactions, the precision of the approach of Lake et al. tended to be lower than the precision of the retrospective approaches (ARE between 1.1 and 1.52). For the method of Trégouet et al., the ARE of tests for G×E interactions varied strongly between the different scenarios and the two risk haplotypes. The values of the ARE calculated for the method of Trégouet et al. relative to the methods of Lin et al. and Chen et al. were in the range of 0 to 1.91 and 0 to 1.77, respectively. Comparable to the ARE values obtained from the analyses of the

scenarios simulated under the null hypothesis, in the scenarios simulated under the alternative hypothesis, the precision of the method of Lin et al. appears to be higher than the precision of the method of Chen et al.

When HWD was specified in the method of Lin et al. (similarly to the analyses of the scenarios simulated under the null hypothesis) the fixation index was estimated to be unequal to zero in less than 1% of the replications of scenarios 2^{dir, het} and 2^{indir, het} but in all replications of scenarios 2^{dir, het} and 2^{indir, het}. Thus, the results of the method of Lin et al. were concordant in scenarios 2^{dir, het} and 2^{indir, het} irrespective of assumption regarding HWE. However, in scenarios 2^{dir, het} and 2^{indir, het}, the power of the method was higher in the analyses not assuming HWD.

Fig. 3. Power of single-marker logistic regression (solid lines) and the four haplotype-based methods (dots) with respect to the G×E interaction effects for the scenarios simulated under a dominant mode of inheritance: 1^d, dir and 1^d, indir (HWE) and 2^d, dir, het, 2^d, indir, het, 2^d, dir, hom, and 2^d, indir, hom (HWD). A: prospective method Lake et al.; B: prospective method of Trégouet et al.; C: retrospective method of Lin et al. (not accounting for HWD); D: retrospective method of Lin et al. (accounting for HWD), and E: retrospective method of Chen et al.

Discussion

We compared the FWER, power, bias, variance, CP, and ARE of four different haplotype-based methods for the analysis of G×E interactions in case-control studies (the prospective approaches of Lake et al. and Trégouet et al., and the retrospective approaches of Lin et al. and Chen et al.) in simulation scenarios for direct and indirect association. Furthermore, the results were compared to the findings of single-marker LR.

In this manuscript, FWER and power were calculated from p values corrected for multiple hypotheses testing using the Bonferroni method, which is known to be conservative. Thus, when investigating the FWER, both the LR and the haplotype-based methods tended to be con-

servative. However, without any correction for multiple hypotheses testing, Lake et al. (for G×E interaction effects only), Trégouet et al. (for main effects only), Lin et al., and Chen et al. (each for main and G×E interaction effects) reported that their methods generally preserved the nominal significance level when haplotype pairs were simulated under HWE. Satten and Epstein compared the type I error and the power of two prospective methods [28, 29] with their own retrospective approach [17] for the analysis of haplotype main effects [19]. With respect to FWER, our results confirm the conclusion arrived at by Satten and Epstein that retrospective approaches are more sensitive to deviations from HWE. Under excess of homozygote haplotype pairs, our results demonstrate that the FWER of the retrospective approaches was high-

er as compared to the HWE scenarios. Thus, use of a less conservative correction for multiple testing can be expected to result in liberal FWERs for the retrospective methods.

Under the alternative hypothesis, in the direct scenarios simulated under a multiplicative mode of inheritance and HWE, we found that LR as well as the retrospective methods of Lin et al. and Chen et al. had greater power to detect interaction effects than the prospective methods of Lake et al. and Trégouet et al. Thereby, the power of the method of Lin et al. to detect interactions tended to be even higher than the power of LR and the method of Chen et al.

To investigate the performance of the methods when the mode of inheritance was wrongly specified, we also analyzed the power of the methods in scenarios simulated under a dominant mode of inheritance but assuming a multiplicative mode of inheritance (see online suppl. tables 9 to 14). We observed that the power was generally lower compared to the scenarios simulated under a multiplicative mode of inheritance. This may firstly be due to the model misspecification, since a multiplicative mode of inheritance was assumed in the analyses [30]. Secondly, whereas the effect size ascribed to heterozygous genotypes at the risk variant is concordant under dominant and multiplicative modes of inheritance, the effect size associated with genotypes that are homozygous for the risk allele is, under a multiplicative mode of inheritance, the square of the corresponding effect size under a dominant mode of inheritance. Thus, the effect of a homozygous genotype is assumed to be stronger under a multiplicative compared to a dominant mode of inheritance. This results in lower power when the disease follows a dominant compared to a multiplicative mode of inheritance. Furthermore, we observed that when the disease was simulated under a dominant mode of inheritance, the retrospective methods of Lin et al. and Chen et al. had generally more power to detect $G \times E$ interaction effects than the prospective methods of Lake et al. and Trégouet et al. Likewise, for haplotype main effects, Satten and Epstein [19] observed that their retrospective approach was more powerful than the investigated prospective approaches under a dominant mode of inheritance. Although we misspecified the mode of inheritance when analyzing these data, the retrospective methods remained superior to the prospective methods despite the fact that the retrospective methods are generally expected to be more sensitive to deviations from modeling assumptions.

In indirect association studies, our results confirm that the power of single-marker approaches can be heavily reduced if LD between true disease variant and genetic marker(s) is only low to moderate [31, 32]. Previous studies have assessed the power of haplotype-based methods using direct association approaches only [8–13, 19]. Therefore, the presented results regarding the performance of the investigated haplotype-based methods in the indirect association scenarios cannot be compared with any other report. The observed decrease in power in the indirect versus direct scenarios was not unexpected, since none of the non-risk variants that constitute the haplotypes in the indirect scenarios was in complete LD with the risk variant (online suppl. tables 1 and 2). However, in the indirect scenarios that were simulated under a multiplicative mode of inheritance, the power of the retrospective methods of Lin et al. and Chen et al. was frequently observed to be higher than the power of the single-marker approach. This was specifically true for the detection of the main and the corresponding $G \times E$ interaction effect of haplotype h_4 , which was uniquely identifiable in both the direct and the indirect scenarios. Furthermore, the retrospective methods of Lin et al. and Chen et al. had generally higher power to detect $G \times E$ interactions than the prospective methods of Lake et al. and Trégouet et al. in the indirect scenarios.

When excess heterozygosity was simulated in the presented analyses, power was generally reduced compared to the HWE scenarios, since haplotype ambiguity and thus the error in haplotype frequency estimates was greater [33] and also the model was misspecified (i.e., HWE was assumed). Similar results were observed by Lake et al. [8] for their prospective method. When excess homozygosity was simulated, even though haplotype ambiguity was decreased, we found that the power only increased for the detection of haplotype h_4 and the corresponding interaction term using the retrospective method of Lin et al. Under HWD, this method was found to be generally more powerful than the other haplotype-based approaches. We did not observe an improved performance of the method of Lin et al. when this approach accounted for HWD. However, this was expected since the model Lin et al. used to account for HWD does not correspond to the model we employed to simulate HWD.

The low CP throughout the different simulation scenarios and methods (with the exception of the method of Lin et al.) reflects that the size of the data sets was too small to obtain good coverage of the 95% confidence intervals even with gold standard methods such as LR.

Nevertheless, good coverage was still obtained with the method of Lin et al. despite the small sample sizes.

The ARE is a measure of the precision of one method relative to another method, which is not linked to the size of the analyzed sample. For the analysis of $G \times E$ interactions, we observed that the precision of the prospective method of Lake et al. was clearly lower than the precision of the retrospective methods, i.e., for the analyses with the method of Lake et al., sample size needs to be increased to obtain the same precision as was achieved with the retrospective methods. The precision of the method of Trégouet et al. was nevertheless superior to the precision of the retrospective approaches in some scenarios simulated under the alternative hypothesis.

The performance of the four haplotype-based approaches may be affected by different factors. Firstly, the results of the method of Chen et al. and Trégouet et al. may be influenced by the limited convergence properties compared to the methods of Lake et al. and Lin et al. Secondly, the approach used to estimate haplotype frequencies may exert influence on the performance of the methods. The prospective methods of Lake et al. and Trégouet et al. estimate haplotype frequencies for the whole sample. These methods do not account for the case-control sampling design and thus are likely to overestimate the frequency of the risk haplotypes since risk haplotypes are overrepresented in the cases. This may result in a loss of power and considerable bias under the alternative hypothesis [30, 34, 35]. We observed significant differences between the certainty of haplotype estimation of prospective versus retrospective methods in some scenarios. These differences could to some extent explain the relatively low power of the investigated prospective methods as compared to the retrospective methods. However, the power of the prospective method of Lake et al. in the analysis of case-control data can be improved by giving different weights to cases and controls to account for the retrospective sampling design. Thirdly, the choice of the reference category may influence the FWER and power of a method. Whereas in the retrospective method of Lin et al. a genetic main/ $G \times E$ interaction effect was tested using all other haplotypes as reference, the reference group of the other methods only comprised a single haplotype (and potentially the rare haplotypes with frequencies ≤ 0.05). Thus, the high power of the retrospective method of Lin et al. may partly be due to the different choice of the reference group as compared to the other approaches. Finally, the superior performance of the retrospective approaches with respect to the power to identify $G \times E$ interactions may also be attributable to the rare

disease and gene-environment independence assumptions [36]. If these assumptions were not valid in the underlying data, the performance of the retrospective methods may be notably deteriorated. However, Lin and Zeng [12] showed that their method may still work well if the rare disease assumption does not hold.

In this paper, we compared the FWER of the four methods using scenarios simulated under the null hypothesis of no association. The power of the approaches was evaluated in scenarios simulated under the alternative hypothesis by considering risk haplotypes only. We did not examine the performance of the different methods with respect to false positive findings (i.e., significant non-risk haplotypes) in the scenarios simulated under the alternative hypothesis. Therefore, future work will address this issue.

In summary, the retrospective methods of Lin et al. and Chen et al. provided greater power to detect $G \times E$ interactions than the methods that do not account for the retrospective sampling design (i.e., the methods of Lake et al. and Trégouet et al.) even if the assumptions of HWE or a multiplicative mode of inheritance were violated. Furthermore, the precision of these methods was superior in comparison to the approach of Lake et al. Moreover, the power of the method of Lin et al. to detect interactions tended to be even higher than the power of single-marker LR. In contrast to the other approaches, this method yielded good coverage of the confidence interval for $G \times E$ interaction effects. However, under HWD, the retrospective methods may not strictly abide by the nominal significance level. The superior performance of the retrospective methods with respect to power may be ascribed to various factors. Firstly, these methods account for the sampling design of the case-control data. Secondly, the efficiency of the retrospective methods is greater, since they exploit the rare disease and gene-environment independence assumptions. In conclusion, for the haplotype-based analysis of $G \times E$ interactions in case-control data, retrospective methods may be superior to methods that do not account for the retrospective sampling design.

Acknowledgements

We thank the reviewers for their constructive comments. This work is part of the subproject 'Statistical methods for the investigation of gene-environment interactions' (PGE-S30T09) of SMP-GEM (NGFN 2), funded by the BMBF in collaboration with the University of Marburg, Germany. Furthermore, this work was supported by the Deutsche Forschungsgemeinschaft Grant BE 3906/2-1 (LB).

References

- Hirschhorn JN: Genetic approaches to studying common diseases and complex traits. *Pediatr Res* 2005;57:74R–77R.
- Akey J, Jin L, Xiong M: Haplotypes vs single marker linkage disequilibrium tests: what do we gain? *Eur J Hum Genet* 2001;9:291–300.
- Clark AG: Inference of haplotypes from PCR-amplified samples of diploid populations. *Mol Biol Evol* 1990;7:111–122.
- Dempster AP, Laird NM, Rubin DB: Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc* 1977;39:1–38.
- Excoffier L, Slatkin M: Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* 1995;12:921–927.
- Stephens M, Smith NJ, Donnelly P: A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001;68:978–989.
- Niu T, Qin ZS, Xu X, Liu JS: Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms. *Am J Hum Genet* 2002;70:157–169.
- Lake SL, Lyon H, Tantisira K, Silverman EK, Weiss ST, Laird NM, Schaid DJ: Estimation and tests of haplotype-environment interaction when linkage phase is ambiguous. *Hum Hered* 2003;55:56–65.
- Tregouet DA, Barbaux S, Escolano S, Tahri N, Golmard JL, Tired L, Cambien F: Specific haplotypes of the P-selectin gene are associated with myocardial infarction. *Hum Mol Genet* 2002;11:2015–2023.
- Tregouet DA, Escolano S, Tired L, Mallet A, Golmard JL: A new algorithm for haplotype-based association analysis: the Stochastic-EM algorithm. *Ann Hum Genet* 2004;68:165–177.
- Lin DY, Zeng D, Millikan R: Maximum likelihood estimation of haplotype effects and haplotype-environment interactions in association studies. *Genet Epidemiol* 2005;29:299–312.
- Lin DY, Zeng D: Likelihood-based inference on haplotype effects in genetic association studies. *J Am Stat Assoc* 2006;101:89–118.
- Chen YH, Chatterjee N, Carroll RJ: Retrospective analysis of haplotype-based case control studies under a flexible model for gene environment association. *Biostatistics* 2008;9:81–99.
- Prentice RL, Pyke R: Logistic disease incidence models and case-control studies. *Biometrika* 1979;66:403–411.
- Rabinowitz D: A note on efficient estimation from case-control data. *Biometrika* 1997;84:486–488.
- Breslow NE, Robins JM, Wellner JA: On the semi-parametric efficiency of logistic regression under case-control sampling. *Bernoulli* 2000;6:447–455.
- Epstein MP, Satten GA: Inference on haplotype effects in case-control studies using unphased genotype data. *Am J Hum Genet* 2003;73:1316–1329.
- Chatterjee N, Carroll RJ: Semiparametric maximum likelihood estimation exploiting gene-environment independence in case-control studies. *Biometrika* 2005;92:399–418.
- Satten GA, Epstein MP: Comparison of prospective and retrospective methods for haplotype inference in case-control studies. *Genet Epidemiol* 2004;27:192–201.
- Spinka C, Carroll RJ, Chatterjee N: Analysis of case-control studies of genetic and environmental factors with missing genetic information and haplotype-phase ambiguity. *Genet Epidemiol* 2005;29:108–127.
- Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA: Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 2002;70:425–434.
- R Development Core Team: R: A Language and Environment for Statistical Computing. 2005. Vienna, Austria, R Foundation for Statistical Computing. Ref Type: Computer Program.
- Sinnwell JP, Schaid DJ, Yu Z. *haplo.stats: Statistical Analysis of Haplotypes with Traits and Covariates when Linkage Phase is Ambiguous*. 2007. R package version 1.3.1. Ref Type: Computer Program.
- Diebolt J, Ip EHS: Stochastic EM: method and application; in Gilks WR, Richardson S, Spiegelhalter DJ (eds): *Markov Chain Monte Carlo in practice*. London, Chapman and Hall, 1996.
- Ip EHS: *Using the Stochastic EM Estimator in the Presence of Missing Data – Theory and Applications*. Department of Biostatistics, Stanford University 1994.
- Tregouet DA, Garelle V: A new JAVA interface implementation of THESIAS: testing haplotype effects in association studies. *Bioinformatics* 2007;23:1038–1039.
- Fallin D, Schork NJ: Accuracy of haplotype frequency estimation for biallelic loci, via the expectation-maximization algorithm for unphased diploid genotype data. *Am J Hum Genet* 2000;67:947–959.
- Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA: Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 2002;70:425–434.
- Zhao LP, Li SS, Khalid N: A method for the assessment of disease associations with single-nucleotide polymorphism haplotypes and environmental variables in case-control studies. *Am J Hum Genet* 2003;72:1231–1250.
- Cordell HJ: Estimation and testing of genotype and haplotype effects in case-control studies: comparison of weighted regression and multiple imputation procedures. *Genet Epidemiol* 2006;30:259–275.
- Zondervan KT, Cardon LR: The complex interplay among factors that influence allelic association. *Nat Rev Genet* 2004;5:89–100.
- Hein R, Beckmann L, Chang-Claude J: Sample size requirements for indirect association studies of gene-environment interactions (G x E). *Genet Epidemiol* 2008;32:235–245.
- Fallin D, Schork NJ: Accuracy of haplotype frequency estimation for biallelic loci, via the expectation-maximization algorithm for unphased diploid genotype data. *Am J Hum Genet* 2000;67:947–959.
- Balding DJ: A tutorial on statistical methods for population association studies. *Nat Rev Genet* 2006;7:781–791.
- Stram DO, Leigh PC, Bretsky P, Freedman M, Hirschhorn JN, Altshuler D, Kolonel LN, Henderson BE, Thomas DC: Modeling and E-M estimation of haplotype-specific relative risks from genotype data for a case-control study of unrelated individuals. *Hum Hered* 2003;55:179–190.
- Piegorsch WW, Weinberg CR, Taylor JA: Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. *Stat Med* 30-1-1994;13:153–162.