

# Rank Transformation in Haseman-Elston Regression Using Scores for Location-Scale Alternatives

Daniel Gerhard<sup>a</sup> Ludwig A. Hothorn<sup>a, b</sup>

<sup>a</sup>Institute of Biostatistics, Leibniz University Hannover, Hannover, Germany; <sup>b</sup>Scripps Genomic Medicine, STSI, Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, Calif., USA

## Key Words

Haseman-Elston regression · Nonparametric · Location-scale alternative

## Abstract

The Haseman-Elston method is a simple regression approach for detecting genetic linkage to quantitative traits in sib-pair studies. Although this method and especially the new extended Haseman-Elston approach are quite robust, there might be some loss of power for non-normally distributed traits. We propose using rank transformation techniques, which either combine the information on a trend in locations and in scales or detect a trend only for a subset of the trait variables for genetically different sibs under linkage. As this rank transformation is based on linear regression, no exact grouping of identity by descent proportions has to be assumed. Simulation results indicate a gain in power compared to recently suggested nonparametric methods.

Copyright © 2009 S. Karger AG, Basel

## Introduction

In sib-pair studies, the Haseman-Elston regression (HE) [1] is a popular method to detect genetic linkage between a locus and a phenotypic trait due to its straightforward and fast computation. It regresses the squared phe-

notypic difference of two sibs over the proportion of alleles shared identity by descent (IBD). Knowing exactly which alleles are identically inherited by the sibs, the proportion of alleles IBD can take only one of three values (0, 0.5, 1), depending on receiving none, one, or two alleles from their parents. For real applications this proportion has to be estimated from incomplete genetic marker data; therefore any value between 0 and 1 may occur with clustering near 0, 0.5, and 1. This problem also arises from calculation of linkage tests for stepwise intervals (in cM) over a whole chromosome, as IBD proportions have to be estimated from only a limited amount of available marker data.

At linkage, there should be some trait difference between the phenotypes for sibs sharing no alleles IBD at a specific locus. Almost no effect of a trait of interest may be found for completely similar genotypes. Further, the variances of the squared phenotypic differences should increase with a decreasing number of shared alleles, as the variance is higher for dissimilar genotypes than for genotypes that are entirely equal.

Until now, many attempts have been made to improve the original HE method, for example, including further information on the mean corrected squared phenotypic sum (Haseman-Elston revisited [2]), and weighting the phenotypic sum and difference by the inverse of their variances (weighted Haseman-Elston regression (wHE) [3]). Although wHE is assumed to be quite robust against

deviations from normality, a loss of power might be possible where the assumptions for a linear regression model are seriously violated. A nonparametric alternative was proposed [4], where a Lepage-type combination [5] of the Jonckheere-Terpstra [6] (JT) and Siegel-Tukey [7] (ST) test statistics was used to merge information about a trend in location and a trend in scale. Simulation showed a gain in power of their method over the wHE approach for non-normal traits, and quite accurate properties under normality.

Our objective is to adapt the HE regression to non-normally distributed phenotypes, keeping the power of the original method for normally distributed residual errors. As in [4], rank transformation scores are chosen to reflect the characteristics of the phenotypes at linkage, but without the constraint of observing the IBD proportions for only the three fixed groups 0, 0.5, and 1.

## Methods

### Haseman-Elston Regression

For a set of independent sib pairs let  $x_{ij}$  be a phenotypic trait of a sib  $i = 1, 2$  in pair  $j = 1, \dots, N$ , assuming the genetic model  $x_{ij} = \mu + g_{ij} + e_{ij}$ , where  $\mu$  is the overall mean, and  $g$  and  $e$  are the genetic and environmental effects. Additionally, let  $Y_j = (x_{j1} - x_{j2})^2$  be the squared phenotypic difference and let  $\pi_j$  be the proportion of alleles IBD. According to Haseman and Elston [1], a simple linear regression can be performed to associate the trait values on the proportion of alleles IBD by

$$E [Y_j | \pi_j] = \alpha + \beta \pi_j, \quad (1)$$

where  $\alpha = \sigma_e^2 + \sigma_g^2$  and  $\beta = -2\sigma_g^2$ . A significantly decreasing slope  $\beta$  indicates linkage at the observed locus. If a genetic effect is present, the distribution of  $Y_j$  conditional on  $\pi_j$  is a mixture of up to three different distributions shifted by genetic effects according to the genotype combinations of the sib pairs. For  $\pi = 0$ , up to nine different distributions occur and at  $\pi = 1$ , one single distribution occurs for the single sibs, illustrated in figure 1, where simulated realizations of normally distributed phenotypes are shown for  $\pi \in \{0, 0.5, 1\}$  in different genetic models. The blue points represent the sib phenotypes, where both sibs own the same allele combination AA-AA, aa-aa, or Aa-Aa. As the sibs are completely identical at this locus, they will only be affected by environmental variance. The genotype combinations AA-Aa, Aa-AA, Aa-aa, and aa-Aa represent sib pairs with one allele IBD, assigned to a proportion of IBD  $\pi = 0$  or 0.5. At linkage, the corresponding phenotypes (red points) are affected by genetic variance in addition to the environmental variance. Also, the phenotypes for the completely different genotype combinations AA-aa and aa-AA at  $\pi = 0$  (green points) are influenced by the genetic variance to a greater extent in additive genetic models. Because only the phenotypes of sib pairs that correspond to different genotypes are showing increasing genetic effects with an increasing number of alleles IBD, linkage can be detected by linear regression over the squared phenotypic difference. But since phenotype combinations for

completely similar genotypes occur over the whole range of IBD proportions, the assumptions for a linear regression are violated in the alternative.

By only looking at the squared phenotypic trait distance  $D_j = -1/2(x_{j1} - x_{j2})^2$ , all information about each sib pair is not used; therefore it may be preferable to incorporate the mean corrected squared phenotypic sum  $S_j = -1/2(x_{j1} + x_{j2} - 2\mu)^2$ .  $D_j$  and  $S_j$  are uncorrelated, but do not share the same variances. To combine both of them, they can be weighted by the inverse of their variances [3]. If

$$w = \frac{\text{var}(S)}{\text{var}(D) + \text{var}(S)},$$

then

$$E [wD_j + (1 - w)S_j | \pi_j] = \alpha + \beta \pi_j. \quad (2)$$

There are many recommendations for estimating the weights and mean corrections. In the following, wHE is defined by estimating the variances directly from the data and taking the sib pair mean as  $\mu$ .

### The Nonparametric Approach by Kim et al. [4]

If the phenotypes of each sib are not normally distributed, the linear regression is not adequate for detecting linkage. Therefore, [4] used the nonparametric trend statistic of Jonckheere [6] (JT) to detect a trend in locations of phenotype groups ordered by the alleles inherited IBD for grouped  $\pi_k$  with  $k = 0, 1$ , and 2, assuming fully informative markers. As the JT test is based on ranks, it is robust against deviations from normality. The JT test can be written as combinations of Mann-Whitney statistics  $U_{kk'}$  by

$$J = \sum_{k=0}^1 \sum_{k'=k+1}^2 U_{kk'}. \quad (3)$$

where  $k$  and  $k'$  denote different groups of sib pairs sharing the same IBD probability. A test of linkage can be performed by a large sample approximation using the standard normal statistic

$$Z_{JT} = \frac{J - E_0(J)}{\sqrt{\text{var}_0(J)}}, \quad (4)$$

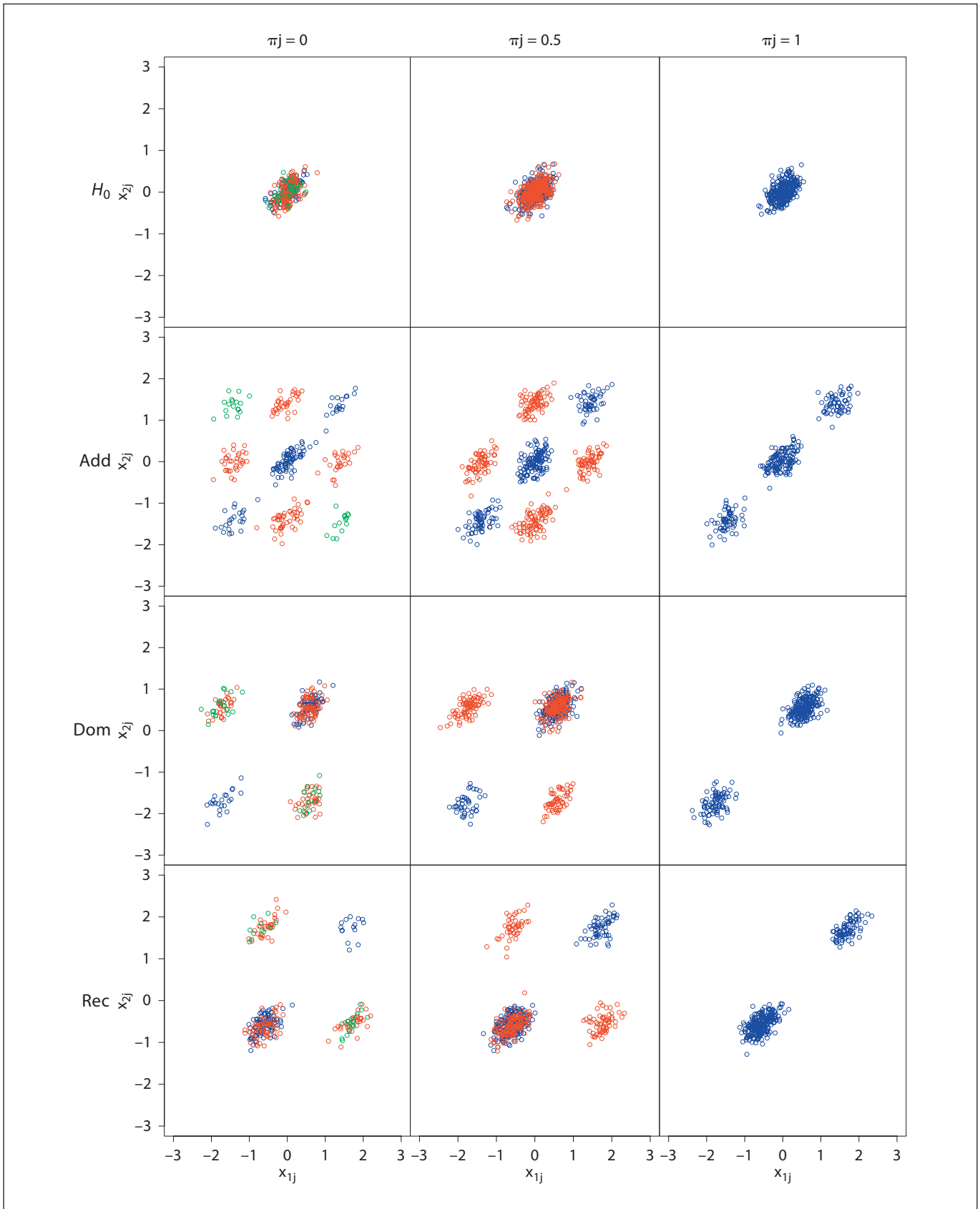
with

$$E_0(J) = \frac{N^2 - \sum_{j=0}^2 n_k^2}{4}$$

and

$$\text{var}_0(J) = \frac{N^2(2N+3) - \sum_{k=0}^2 n_k^2(2n_k+3)}{72}.$$

**Fig. 1.** Scatterplot for random normal phenotypes of the two sibs in the  $j$ -th sib pair for  $h^2 = 0$  ( $H_0$ ) and  $h^2 = 0.5$  for an additive (Add), dominant (Dom), and recessive (Rec) mode of inheritance with  $p = 0.5$ ,  $N = 1,000$ ,  $\sigma_e^2 = 0.05$  and an overall sib correlation of  $\rho = 0.5$ . The blue-colored points represent the phenotypes of the sib genotype combinations AA-AA, aa-aa, Aa-Aa; the red points display AA-Aa, Aa-AA, Aa-aa, aa-Aa and the green points mark the complete allele differences AA-aa and aa-AA.



As the groups IBD also show differences in variances under the alternative, a test for differences in scales might be suitable to detect a quantitative trait loci (QTL). Kim et al. [4] chose the Siegel-Tukey test [7] (ST) for this kind of scale hypotheses. Because this test compares only two sample variances, they plugged the Siegel-Tukey scores into the JT statistic to generalize it for a trend test ( $Z_{ST}$ ). The two test statistics for location and for scale have to be combined to test for both trends, using the Lepage approach [5]

$$Z_{JT+ST} = \frac{Z_{JT} + Z_{ST}}{\sqrt{2}} \quad (5)$$

for a combination, which assumes approximate normality and uncorrelated test statistics.

#### Regression-Type Rank Transformation Combining Scores for Location and Scale Alternatives

Rank transformation is a common way to deal with unknown distributions in various applications [8, 9]. In the following,  $R(Y_j)$  denotes the rank of every  $Y_j$  over the space from 1 to  $N$ , assigning average ranks in the case of ties. First, we want to test for a significant trend in locations; therefore Wilcoxon scores [10] are applied to the response variable, so that

$$a(R) = \frac{R}{N+1},$$

where the lowest score is assigned to the lowest ranked value and any subsequent score increases, according to the ranking. Taking the scores as response, the linear regression model is written as

$$E(a(R_j) | \pi_j) = \alpha + \beta \pi_j. \quad (6)$$

The test for the slope parameter  $H_0: \beta \geq 0$  and  $H_A: \beta < 0$  is performed similarly to the original Haseman-Elston method

$$T = \frac{\hat{\beta}}{\sqrt{\widehat{\text{var}}(\hat{\beta})}}. \quad (7)$$

$T$  can be assumed to be asymptotically  $t$  distributed with  $N - 2$  degrees of freedom, providing the calculation of a p value  $p_{loc}$ . Under equally spaced  $\pi_j$ , this test is equivalent to the test for correlation based on Spearman's  $\rho$  [9], which is also proposed by Haseman and Elston [1] for nonparametric linkage analysis. Here, we assume incomplete marker data and hence randomly spaced  $\pi_j$ .

To include additional information about a trend in scales to support the hypotheses of a linked locus, we have to choose appropriate rank scores and plug these into the regression model as response variables as well. Ansari-Bradley scores [11] can be used for the scale regression:

$$b(R) = \frac{N+1}{2} \left| R - \frac{N+1}{2} \right|,$$

where small scores are assigned to both extremes, with an increasing score for the ranks in the middle. A test for the slope parameter in the regression model

$$E(b(R_j) | \pi_j) = \alpha + \beta \pi_j \quad (8)$$

results in a corresponding p value  $p_{sca}$ . In Kim et al. [4], Siegel-Tukey [7] scores are favored over the Ansari-Bradley scores, as the Siegel-Tukey method avoids giving the same scores for small and

large ranks; nevertheless we could not find large differences between the two methods, preferring the Ansari-Bradley scores for their straightforward software implementation. For a combination of both nonparametric tests for location and scale, the two rank scores have to be uncorrelated. The covariance between score statistics is zero for the combination of an odd with an even translation invariant statistic [12]. It is easy to show that for the Wilcoxon (odd:  $a(R) = -a(N+1-R)$ ) and the Ansari-Bradley scores (even:  $b(R) = b(N+1-R)$ ) this criterion is fulfilled, whereas the joint test decision might be biased under the alternative, since the regression with location scores assumes the homogeneity of scales and vice versa. A method to combine both regression outcomes, for location and for scale, is Fisher's combination test [13] (Fis), which uses the logarithm of two p values. A test statistic can be defined as

$$T_{Fisher} = -2(\log(p_{loc}) + \log(p_{sca})), \quad (9)$$

with  $T_{Fisher}$  being  $\chi^2$ -distributed with two times the number of combined p values as degrees of freedom (here d.f. = 4). For a combination of just two uncorrelated p values, this combination test was shown to be uniformly most powerful [14].

#### Detecting Linkage for A Subset of $Y_j$

A subset of sib pairs over all IBD groups share the same distribution, whereas differences occur with varying IBD probabilities under linkage only for some of them. In the context of clinical drug research, when only a subset of patients can be expected to respond to a treatment, Conover and Salsburg [15] proposed to model this shift in distributions as a Lehmann [16] alternative. They suggested the approximate scores

$$s(R) = \left( \frac{R}{N+1} \right)^{\theta-1}$$

with  $\theta = 5$  based on empirical results. By plugging these scores into the linear model of the Haseman and Elston approach,

$$E(s(R_j) | \pi_j) = \alpha + \beta \pi_j, \quad (10)$$

a significant test for  $\hat{\beta}$  will correspond to a trend for only a subset of  $Y_j$  (CS). Thus, an attempt is made to distinguish between two groups of phenotype differences: an uninformative group, where sib pairs possess similar genotypes and the corresponding phenotypes are only affected by environmental variability, and a group consisting of genetically different sibs under linkage with phenotypes being additionally influenced by genetic variability. Assuming a mixed distribution based on these two groups may result in a higher power compared with the investigation of a trend for the whole data, then distinguishing between a group of sib pairs under linkage and an uninformative group of sib pairs with similar genotypes is quite a good approximation to the various response distributions of the trait differences under the alternative hypothesis. For large sample sizes this fact need not be considered. As the Conover-Salsburg scores are a weighted variant of Wilcoxon scores ( $\theta = 2$ ), giving more weight to higher ranked values, they are used to detect a trend in locations at linkage. With only a subgroup of observations showing a trend in locations under the alternative and the others remaining under the null hypothesis of no trend, a trend in scales is thereby indirectly assumed. Thus, the Conover-Salsburg method represents a kind of compromise between location and scale tests.

**Table 1.** Size and power of tests in parametric and rank-based regression models assuming contaminated normal distribution (2.5% phenotypes with 5 times higher standard deviation)

<i>p</i>	<i>h</i> <sup>2</sup>	Dominant					Additive					Recessive				
		HE	wHE	CS	Fis	Kim	HE	wHE	CS	Fis	Kim	HE	wHE	CS	Fis	Kim
0.1	0	0.052	0.050	0.050	0.051	0.050	0.052	0.050	0.050	0.051	0.050	0.052	0.050	0.050	0.051	0.050
	0.1	0.084	0.137	0.223	0.200	0.203	0.084	0.137	0.218	0.196	0.199	0.081	0.126	0.101	0.092	0.092
	0.3	0.252	0.545	0.834	0.791	0.790	0.253	0.536	0.807	0.758	0.760	0.170	0.300	0.111	0.097	0.098
	0.5	0.618	0.928	0.995	0.991	0.991	0.616	0.916	0.993	0.986	0.987	0.252	0.408	0.113	0.100	0.100
	0.6	0.828	0.987	0.999	0.998	0.999	0.823	0.982	0.999	0.998	0.998	0.277	0.433	0.115	0.099	0.100
0.5	0	0.051	0.051	0.050	0.050	0.049	0.051	0.051	0.050	0.050	0.049	0.051	0.051	0.050	0.050	0.049
	0.1	0.085	0.138	0.221	0.201	0.203	0.085	0.138	0.226	0.204	0.205	0.084	0.137	0.221	0.200	0.204
	0.3	0.250	0.550	0.850	0.819	0.812	0.254	0.556	0.872	0.840	0.828	0.249	0.549	0.851	0.820	0.812
	0.5	0.619	0.938	0.998	0.997	0.997	0.636	0.942	0.999	0.999	0.998	0.620	0.938	0.998	0.997	0.997
	0.6	0.831	0.991	1.000	1.000	1.000	0.847	0.992	1.000	1.000	1.000	0.830	0.991	1.000	1.000	1.000
0.7	0	0.050	0.049	0.051	0.050	0.050	0.050	0.049	0.051	0.050	0.050	0.050	0.049	0.051	0.050	0.050
	0.1	0.085	0.137	0.204	0.184	0.188	0.086	0.138	0.226	0.204	0.207	0.086	0.139	0.228	0.205	0.208
	0.3	0.239	0.519	0.686	0.621	0.640	0.254	0.553	0.865	0.832	0.821	0.252	0.555	0.880	0.856	0.840
	0.5	0.571	0.892	0.895	0.833	0.852	0.633	0.940	0.999	0.999	0.998	0.631	0.946	1.000	0.999	0.999
	0.6	0.770	0.970	0.927	0.868	0.885	0.845	0.992	1.000	1.000	1.000	0.844	0.994	1.000	1.000	1.000

*Simulations*

In order to compare the power of the different methods, we performed simulations under a variety of conditions. The simulation settings were mainly adopted from Kim et al. [4]. First, IBD scores were generated from a trinomial distribution under the assumption of fully informative, biallelic markers with cell probabilities 0.25 for homozygous and 0.5 for heterozygous loci. Then, the trait genotypes for *N* = 500 sib pairs were randomly drawn from a multinomial distribution with cell probabilities conditional to the generated IBD probabilities with allele frequencies of *p* = 0.1, 0.5, 0.7. Phenotypic values for each sib were generated from normal, contaminated normal, log-normal, and Laplace distributions, choosing 1 for the residual variance and varying the percentage of variance explained by the QTL (*h*<sup>2</sup>) for an additive, dominant, or recessive mode of inheritance. For each sib pair, an overall phenotypic residual correlation is chosen at *ρ* = 0.5. For the contaminated normal distribution, 2.5 or 10% of the phenotypes are assumed to have a residual standard deviation five times larger. The Laplace distributed phenotypes with a variance of two are generated, according to Fernandez et al. [17], from two random, normally distributed variates for each sib with a correlation of 0.5454. In addition, the performance of the methods is investigated for different numbers of observations from *N* = 50, 100, ..., 1,000 for an additive model, an allele frequency of *p* = 0.5, and heritability of *h*<sup>2</sup> = 0.2. The simulation was conducted in the statistical software R [26]; every simulation step was repeated with 100,000 runs.

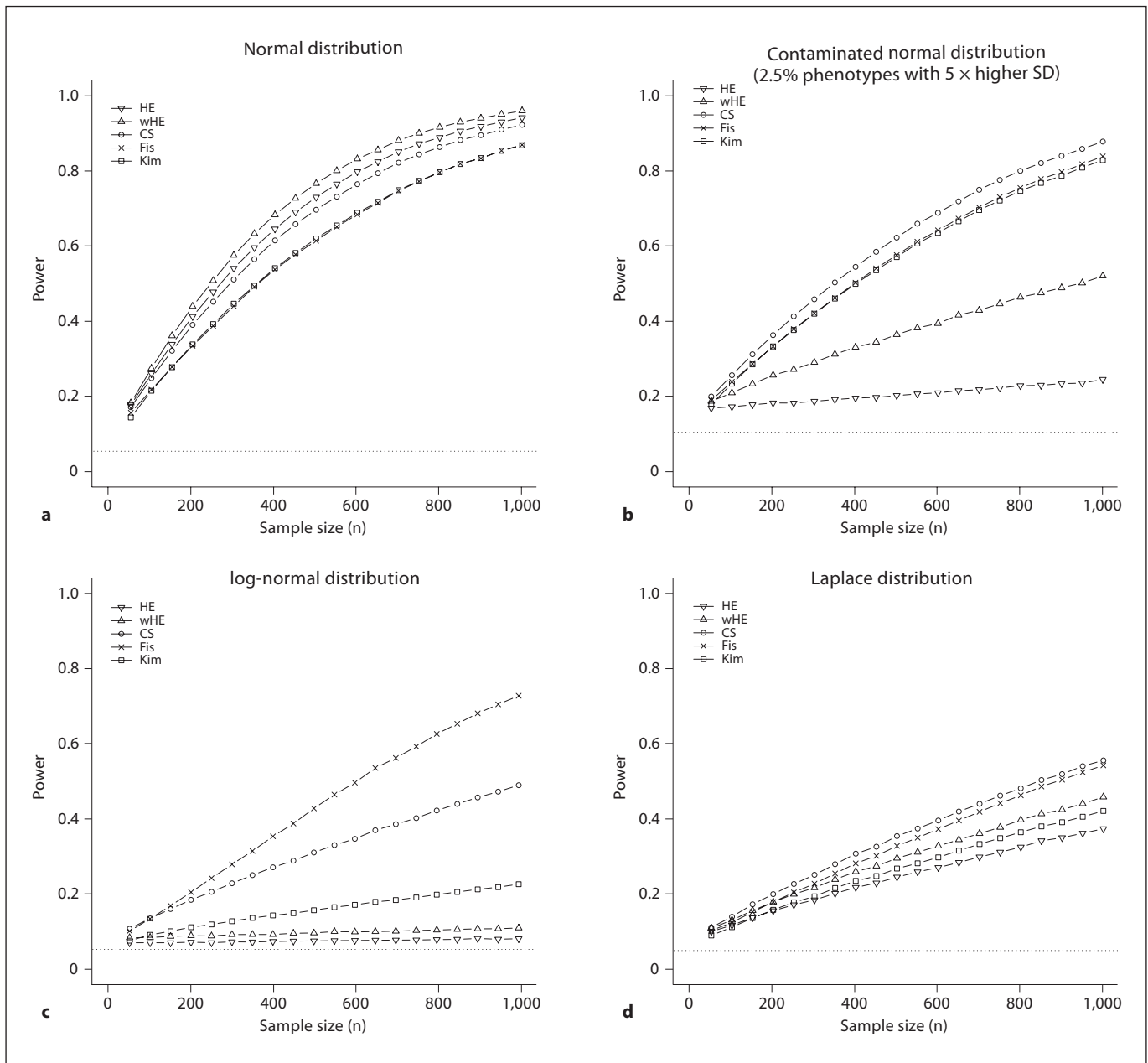
**Results**

The simulation results under the null hypothesis show that all methods hold a nominally level *α* for every investigated residual distribution: as under the null hypothesis

of no significant trend, there is no difference in distributions for any sib pairs. If we look at the power simulations for data, assuming normally distributed errors (for details, see [18]), it is obvious that the wHE method is the most powerful of the methods under investigation, but in almost every situation there is no significant difference in power to the CS approach. Only at small allele frequencies, especially under a recessive mode of inheritance, do the rank transformation methods show a loss in power. One reason might be the detection of the small amount of sib pairs showing an effect as outliers, whereas the parametric methods incorporate this information. Both approaches combining scores for location and scale show quite a similar power.

Figure 2 shows that all rank methods are rather stable, even if the sample size is small. A sample size of 100 sib pairs should be sufficient to assume approximate normality for the rank statistics. When we assume that 2.5% of the phenotypes possess a higher residual variance (table 1), the rank transformation methods are more robust than the wHE method.

Here, the CS approach also appears to have an advantage in power. If the proportion of phenotypes with higher residual variance increases, the power of all approaches decreases, but to a lesser extent for the methods incorporating scale effects (see [18]). Looking at residual distributions with long tails, for example the log-normal distribution (see [18]), the wHE has problems detecting

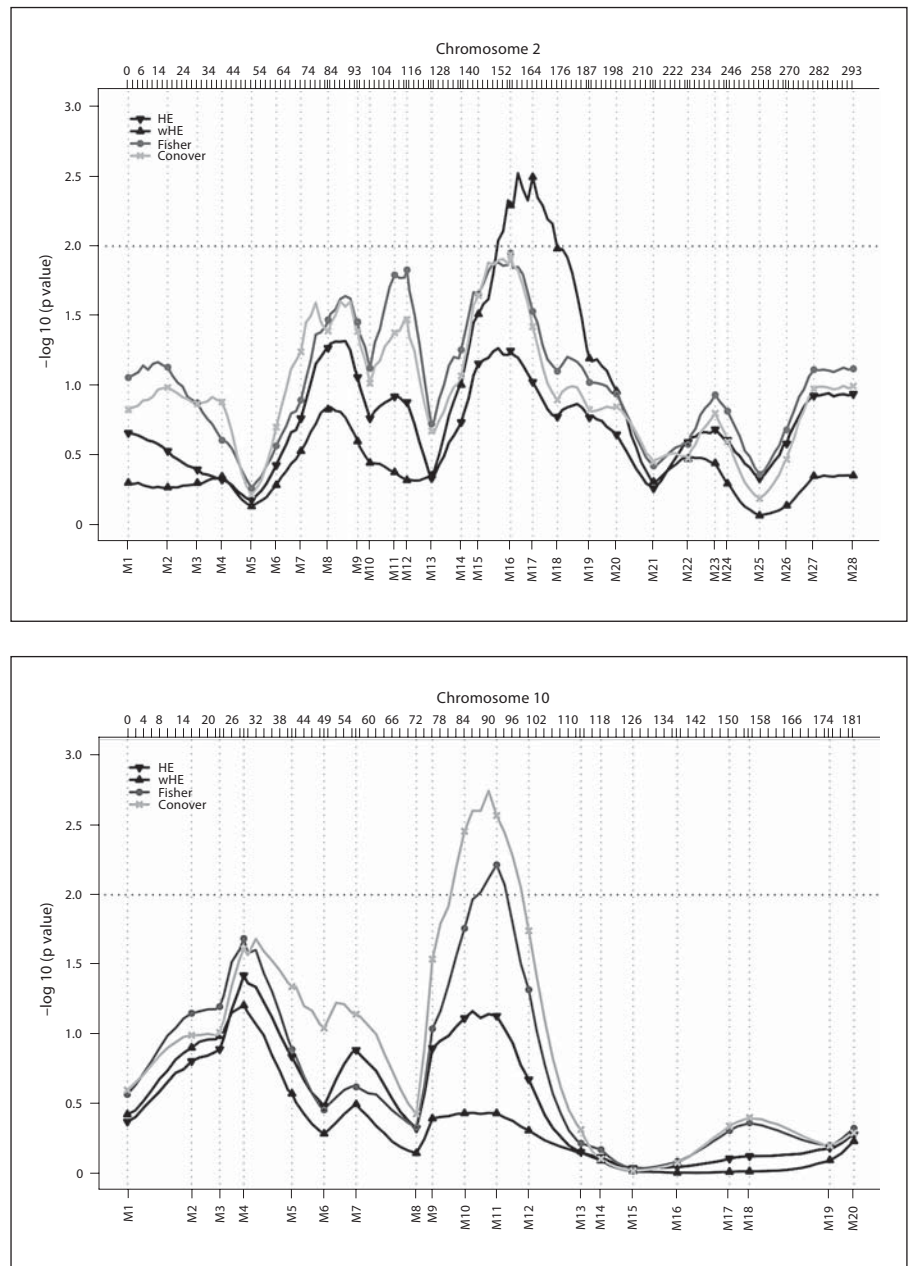


**Fig. 2.** Power for regression model-based tests for various sample size settings for an additive genetic model with  $p = 0.5$ ,  $h^2 = 0.2$ .

any QTL at all. In their simulations, Kim et al. [4] showed a gain in power at small QTL effects for only the location scores, whereas the combination of locations and scales by their proposed Lepage statistic appears to be less powerful. The location-scale combination by Fisher's test does not suffer from including the information about scales and displays almost the same properties as the location statistic.

### Example

As an example, we applied our methods to the Beaver Dam Eye Study [19] with systolic blood pressure as a single trait. Information from a sample of 120 families, some of them containing more than one sib pair, was investigated. As the family size has an influence on the results, proved by choosing random pairs out of each family (not



**Fig. 3.** Interval mapping for chromosomes 2 and 10 for systolic blood pressure.

shown here), we used a bootstrap method for the  $Y_i$  to avoid bias from the correlation of sibs in one family. Multipoint IBD probabilities over intervals of 2 cM were calculated with the software S.A.G.E [25]. Across chromosome 2, the original HE regression and the nonparametric methods follow the same curvature of p-values, where the rank transformation results in overall lower p values (fig. 3). There is one value of the response vector that is remarkably higher than the other measurements, possibly creating some bias, which is not weighted to this ex-

tent by the rank transformation methods. The wHE approach shows a slightly different pattern with a peak from 145–176 cM above  $p = 0.01$ . Wang and Elston [20], who analyzed the same dataset, had previously reported linkage evidence in this region. Across chromosome 10, the nonparametric methods show a high peak at around 90 cM (fig. 3). For the squared trait difference at this special marker, an exceptionally high value occurs at IBD probability 1, which causes the parametric methods to miss the indication of a downward regression trend.

## Discussion

If it cannot be assumed that a trait is normally distributed, researchers may increase power to detect a QTL using rank transformation methods. The approach of combining scores for location and scales by Fisher's combination of *p* values shows almost the same properties, and in some situations a gain in power, as with the Kim et al. [4] Lepage combination. As an alternative, modeling the squared trait distribution as a mixture of two distributions with Conover-Salsburg scores results in a gain in power, especially if only a few values do not fit into the normal assumption. There are many studies demonstrating the limitations of rank transformation in complex designs and multiple regression [21, 22]. They show a severe inflation of the type I error rate and a substantial loss of power. Nevertheless, in our simulations we could not detect any of these violations in our one-way regression setting, but we would not suggest including additional covariates into the model.

The original HE regression takes only families consisting of exactly two sibs and their parents into account. In larger families, or even more complex pedigree structures, the correlation between pedigree members has to be considered. This can be done by variance components methods [23] or by adopted regression-based methods. Sham et al. [24] show the equivalence of their regression method to variance components in general pedigrees with the robustness of the wHE method. Their method is based on a pooled estimate of QTL variances, regressing the squared phenotypic difference and squared phenotypic sum over the proportion of IBD sharing, weighted by sib correlations. In families with only two sibs, this regression method is quite similar to the wHE method in terms of power. With increasing family size, the Sham et al. method gains some power, as it exploits the correlation

between family members. Reproducing this method by a rank transformation approach might be more difficult, as the regression coefficients cannot be interpreted anymore in terms of explained QTL variances or regression coefficients. Furthermore, by observing the transformed ranks of the squared trait difference and sum simultaneously, these may no longer be independent even for families consisting of only two sibs.

The Kim et al. approach is not feasible if there are large amounts of missing marker information. The discreteness, which originates from comparing single ordered groups, makes it impossible to access additional information given by algorithms that provide continuous IBD probabilities. Therefore, the modification of the established regression procedures seems to be applicable to a wider range of problems, for example, to interval mapping.

If IBD probabilities are given, the regression over rank scores can be done with any statistical software capable of fitting a linear model, such as R or the SAS System. The rank transformation may also be easily implemented into any available software providing the classical HE approach. Computer code for the implementation of the approach in R and for the simulations shown here can be obtained from the corresponding author.

## Acknowledgements

Some of the results of this paper were obtained by using the program package S.A.G.E., which is supported by a U.S. Public Health Service Resource Grant (RR03655) from the National Center for Research Resources. Part of the work of LAH was performed during a sabbatical stay at TSRI, La Jolla. I would like to thank the group leader, Dr. Nickolas Schork, for generous local support. We would also like to thank Dr. Andreas Ziegler, Luebeck, and two anonymous reviewers for their helpful comments which contributed greatly to the quality of this paper.

## References

- 1 Haseman JK, Elston RC: The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet* 1972;2:3–19.
- 2 Elston RC, Buxbaum S, Jacobs KB, Olson JM: Haseman and Elston revisited. *Genet Epidemiol* 2000;19:1–17.
- 3 Forrest W: Weighting improves the 'New Haseman-Elston' method. *Hum Hered* 2001; 52:47–54.
- 4 Kim MK, Hong YJ, Song HH: Nonparametric trend statistic incorporating dispersion differences in sib pair linkage for quantitative traits. *Hum Hered* 2006;62:1–11.
- 5 Lepage Y: A combination of Wilcoxon's and Ansari-Bradley's statistics. *Biometrika* 1971; 58:213–217.
- 6 Jonckheere AR: A distribution-free *k*-sample test against ordered alternatives. *Biometrika* 1954;41:133–145.
- 7 Siegel S, Tukey JW: A nonparametric sum of ranks procedure for relative spread in unpaired samples. *J Am Stat Assoc* 1960;55: 429–445.
- 8 Iman RL, Conover WJ: The use of the rank transform in regression. *Technometrics* 1979;21:499–509.



- 9 Hettmansperger TP, McKean JW: Robust Nonparametric Statistical Methods. London, Arnold, 1998.
- 10 Wilcoxon F: Individual comparisons by ranking methods. *Biometrics Bulletin* 1945; 1:80–83.
- 11 Ansari AR, Bradley RA: Rank-sum tests for dispersion. *Annals of Mathematical Statistics* 1960;31:1174–1189.
- 12 Randles RH, Hogg RV: Certain uncorrelated and independent rank statistics. *J Am Stat Assoc* 1971;66:569–574.
- 13 Hedges LV, Olkin I: *Statistical Methods for Meta-Analysis*. Orlando, Academic Press, 1985.
- 14 Van Zwet WR, Oosterhoff J: On combination of independent test statistics. *Annals of Mathematical Statistics* 1967;38:659–680.
- 15 Conover WJ, Salsburg DS: Locally most powerful tests for detecting treatment effects when only a subset of patients can be expected to ‘respond’ to treatment. *Biometrics* 1988;44:189–196.
- 16 Lehmann EL: The power of rank tests. *Annals of Mathematical Statistics* 1953;24:23–43.
- 17 Fernandez JR, Etzel C, Beasley TM, Shete S, Amos CI, Allison DB: Improving the power of sib pair quantitative trait loci detection by phenotype winsorization. *Hum Hered* 2002; 53:59–67.
- 18 Gerhard D: Simulation study comparing nonparametric alternatives to Haseman-Elston regression. Technical Report, Leibniz University of Hannover, 2008.
- 19 Klein R, Klein BE, Jensen SC, Meuer SM: The five-year incidence and progression of age-related maculopathy: the Beaver Dam Eye Study. *Ophthalmology* 1997;104:7–21.
- 20 Wang T, Elston RC: Regression-based multivariate linkage analysis with an application to blood pressure and body mass index. *Ann Hum Genet* 2007;71:96–106.
- 21 Brunner E, Neumann N: Rank tests in  $2 \times 2$  designs. *Statist Neerlandica* 1986;40:251–272.
- 22 Headrick TC, Rotou O: An investigation of the rank transformation in multiple regression. *Comput Stat Data Anal* 2001;38:203–215.
- 23 Amos CI: Robust variance-components approach for assessing genetic linkage in pedigrees. *Am J Hum Genet* 1994;54:535–543.
- 24 Sham PC, Purcell S, Cherny SS, Abecasis GR: Powerful regression-based quantitative-trait linkage analysis of general pedigrees. *Am J Hum Genet* 2002;71:238–253.
- 25 S.A.G.E.: *Statistical Analysis for Genetic Epidemiology*. Release 5.4: <http://genepi.cwru.edu/>.
- 26 R Development Core Team: *R: A language and environment for statistical computing*. Vienna, R Foundation for Statistical Computing, 2007.