Genetic Epidemiology 30: 384–396 (2006)

Using Sex-Averaged Genetic Maps in Multipoint Linkage Analysis When Identity-by-Descent Status is Incompletely Known

Tasha E. Fingerlin,[1]* Gonçalo R. Abecasis[2] and Michael Boehnke[2]

[1]*Department of Preventive Medicine and Biometrics, School of Medicine, University of Colorado at Denver and Health Sciences Center, Denver, Colorado*
[2]*Department of Biostatistics and Center for Statistical Genetics, School of Public Health, University of Michigan, Ann Arbor, Michigan*

The ratio of male and female genetic map distances varies dramatically across the human genome. Despite these sex differences in genetic map distances, most multipoint linkage analyses use sex-averaged genetic maps. We investigated the impact of using a sex-averaged genetic map instead of sex-specific maps for multipoint linkage analysis of affected sibling pairs when identity-by-descent states are incompletely known due to missing parental genotypes and incomplete marker heterozygosity. If either all or no parental genotypes were available, for intermarker distances of 10, 5, and 1 cM, we found no important differences in the expected maximum lod score (EMLOD) or location estimates of the disease locus between analyses that used the sex-averaged map and those that used the true sex-specific maps for female:male genetic map distance ratios 1:10 and 10:1. However, when genotypes for only one parent were available and the recombination rate was higher in females, the EMLOD using the sex-averaged map was inflated compared to the sex-specific map analysis if only mothers were genotyped and deflated if only fathers were genotyped. The inflation of the lod score when only mothers were genotyped led to markedly increased false-positive rates in some cases. The opposite was true when the recombination rate was higher in males; the EMLOD was inflated if only fathers were genotyped, and deflated if only mothers were genotyped. While the effects of missing parental genotypes were mitigated for less extreme cases of missingness, our results suggest that when possible, sex-specific maps should be used in linkage analyses. *Genet. Epidemiol.* 30:384–396, 2006.

Key words: sex-specific maps; linkage analysis

Contract grant sponsor: National Institute of Health Grant; Contract grant numbers: HG00376; HG02651;
Contract grant sponsor: National Institute of Health Training Grant; Contract grant number: HG0040.
*Correspondence to: T.E. Fingerlin, Ph.D. 2400 E. 9th Avenue, Box B-119, Department of Preventive Medicine and Biometrics, University of Colorado at Denver and Health Sciences Center, Denver, CO 80262. E-mail: tasha.fingerlin@uchsc.edu
Received 21 November 2005; Revised 26 January 2006; Accepted 14 February 2006
Published online 9 May 2006 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/gepi.20151

INTRODUCTION

Multipoint linkage analysis is a standard tool in the search for genetic variants that predispose to Mendelian and complex genetic diseases. Multipoint methods are generally more powerful than single-point methods [Lathrop et al., 1984], but typically assume a known genetic map. Misspecification of the genetic map has the potential to compromise the estimation and testing procedures used in multipoint linkage analysis. Specifically, inflation or deflation of the lod score, loss of power to detect linkage or an increase in the false-positive rate, and bias in the disease locus position estimate are possible depending on the type and degree of misspecification [Hauser, 1998; Halpern and Whittemore, 1999; Daw et al., 2000].

Marker map misspecification can arise from uncertainty due to the estimation process (sampling error) or from over-simplified models of the biological process of recombination. The number of meioses, and therefore the number of recombination events, used to infer intermarker distances for genetic maps is often relatively small (<200). As a result, many current genetic maps have large sampling errors [Broman et al., 1998], particularly for dense marker spacing. For example, Daw et al. [2000] estimated that with error-free marker data, the width of the 95% confidence interval for a 10 cM intermarker distance typically ranges from 4 to 19 cM given data on 200 meioses and markers with average heterozygosity .77. Robinson [1996] gave a thorough review of the implications of different types of variation in recombination rates

for building human genetic maps. Most notably for the current paper, evidence for sex variation in recombination rates indicates that sex-averaged maps fail to incorporate relevant sex-specific genetic distance information [NIH/CEPH Collaborative Mapping Group, 1992; Broman et al., 1998].

Despite the evidence that the female:male genetic map distance ratio varies dramatically along chromosomes and across the genome [Straub et al., 1993; Broman et al., 1998; Mohrenweiser et al., 1998; Kong et al., 2002], most multipoint linkage studies use sex-averaged rather than sex-specific genetic maps. This is true for several reasons. First, since the number of meioses used to estimate sex-specific maps is about half the number used to estimate sex-averaged maps, the sampling error necessarily is larger in the sex-specific case. Second, in many studies, little or no parental data are available, particularly when diseases with a late age of onset are studied. As such, it is commonly assumed that use of sex-specific maps may not be advantageous since the outcomes of maternal and paternal meioses are difficult to assess independently. Third, and perhaps most importantly, even if reliable sex-specific maps are available, some widely used analysis packages, including GENEHUNTER [Kruglyak et al., 1996], and historically, Allegro [Gudbjartsson et al., 2000] and Merlin [Abecasis et al., 2002], have not supported sex-specific maps.

Daw et al. [2000] assessed the impact on multipoint linkage analysis of misspecifying the sex-averaged distance or using the sex-averaged distance instead of sex-specific distances assuming fully informative meioses. They found that in the presence of linkage, the expected lod score calculated under the misspecified map was always less than that calculated under the true map, but that the difference was small ($<4\%$) unless both the sex-averaged distance and the female:male map distance ratio were substantially misspecified. In contrast, they found that given no linkage, the expected lod score calculated under a sex-averaged map was always greater than that calculated under sex-specific maps when the female:male map distance ratio was not equal to one. This inflation in the lod score led to modestly increased false-positive rates in some cases.

Here we extend the work of Daw et al. [2000] to the more typical situation of partially informative meioses. In any study, most meioses are not fully informative due to limited marker density and incomplete marker heterozygosity; the problem is more pronounced in studies of affected sibships when parental genotypes are missing. Estimates of allele sharing among relatives are more dependent on the assumed intermarker distances when meioses are not fully informative.

We have implemented modeling of sex-specific recombination fractions in the Merlin pedigree analysis package [Abecasis et al., 2002]. Here, we use Merlin to evaluate the impact of using a sex-averaged map rather than sex-specific maps in multipoint linkage analyses of affected sibships with genotype data for 0, 1, or 2 parents. We evaluate the expected lod score curves, power, and size of non-parametric linkage tests when the correct sex-specific maps are used and when a sex-averaged map is used. In addition, we compare the estimates of the disease location and 1-lod support interval for each setting.

We find that using a sex-averaged map in place of the biologically more plausible sex-specific maps can lead to increased false-positive rates or decreased power if there is a substantial difference in the number of informative male and female meioses and marker density is relatively low (<1 marker per 5 cM). While for most studies, the imbalance between available maternal and paternal genotypes is likely to be modest compared to our most extreme cases, our results suggest that sex-specific maps should be used in linkage analyses.

METHODS

DISEASE MODEL AND DATA

Let $\mathbf{M}_m$ and $\mathbf{M}_f$ be the vectors of the true male and female intermarker map distances and $\tilde{\mathbf{M}}_m$, $\tilde{\mathbf{M}}_f$ those assumed for analysis. Similarly, let $\mathbf{M}$ be the vector of sex-averaged intermarker map distances (each sex-averaged distance is the mean of the true sex-specific distances) and $\tilde{\mathbf{M}}$ that assumed for analysis. Note that in practice, sex-averaged map distances are estimated using all informative meioses simultaneously and are not in general simply the average of the two sex-specific map distances.

We compared via simulation the expected maximum lod scores (EMLOD) when the true sex-specific maps are assumed to that when the sex-averaged map is assumed: $\tilde{\mathbf{M}}_m = \mathbf{M}_m$ and $\tilde{\mathbf{M}}_f = \mathbf{M}_f$ versus $\tilde{\mathbf{M}}_m = \tilde{\mathbf{M}}_f = \mathbf{M}$. In addition, we compared the EMLOD when either misspecified sex-specific $\tilde{\mathbf{M}}_m \neq \mathbf{M}_m$, $\tilde{\mathbf{M}}_f \neq \mathbf{M}_f$ or misspecified

sex-averaged $\tilde{\mathbf{M}} \neq \mathbf{M}$ maps are assumed. For the second set of comparisons we were particularly interested in evaluating the impact of misspecifying the sex-specific map distances, since those maps typically have greater sampling error.

PATTERNS IN RATIO OF FEMALE TO MALE GENETIC MAP DISTANCES

On average across the genome, the ratio of female:male genetic distance is $r \approx 1.6$. Chromosome-specific averages range from 1.2 for chromosomes 15 and 19 to 2.0 for chromosome 8 [Broman et al., 1998]. For all metacentric chromosomes, a peak in the ratio is observed at or near the centromere and can be as large as $r = 11$ [Mohrenweiser et al., 1998], but the exact pattern of the change in the ratio varies across chromosomes.

We considered two different patterns of the female:male map distance ratio across the chromosome (Fig. 1): (a) $r = 10$ at the centromere and decreasing toward the telomeres as for chromosome 19 [Broman et al., 1998], and (b) r variable across the map with two distinct peaks as for chromosome 11 [Broman et al., 1998]. To assess the impact of the disease gene location, d, we also varied the position of the disease locus (Fig. 1). We recognize that there is some evidence for much more variable and extreme values of r over short distances [Straub et al., 1993; Kong et al., 2002]. We chose to use these rather smooth functions of r rather than extreme cases for simplicity.

SIMULATION STUDY

We carried out computer simulations to compare the expected lod scores, false-positive rates, and power for analyses using sex-specific genetic maps to those using sex-averaged maps. We initially simulated autosomal marker data for 500 ASPs and their parents for a single, additive locus with effect size $\lambda_s = 1$, 1.1, 1.25, and 2. To investigate the impact of (a) the number of ASPs, (b) sibship size, and (c) pedigree structure, we also simulated autosomal marker data for (a) 250 and 1000 ASPs, (b) for affected sibships of size 3, 4, or 5, and (c) for families with either an affected half-sib or affected first cousin pair and their parents. To keep the total number of genotypes approximately constant, we simulated marker data for 333, 250, and 200 nuclear families for sibship sizes 3, 4, and 5, respectively. For affected sibships with more than two siblings and more extended pedigrees, we fixed the disease-allele frequency at .20 and penetrances for 0, 1, and 2 copies of the disease allele at .050, .175, and .300, respectively, which corresponds to $\lambda_s = 1.25$ for ASPs. We generated parental genotypes based on the allele frequencies of the markers assuming Hardy-Weinberg and linkage equilibria. We generated genotypes for the affected sibs conditional on the genotypes of the parents and model parameters. We determined whether each individual was affected on the basis of the penetrance function for his/her genotype and kept only those sibships with all siblings affected. We simulated data for a map of markers (Fig. 1 and see above), each with four equally frequent alleles, equally spaced at 1 cM (sex-averaged distance) intervals. For sib pairs only, we also simulated data for a map of SNP markers with allele frequencies = .50 when $\lambda_s = 1.0$ and 1.25. We used Haldane's no-interference mapping function [1919] to convert the map distances to recombination fractions for analysis.

For each simulation condition with $\lambda_s = 1$ and > 1, we generated 10,000 and 5,000 replicate data sets, respectively. For each replicate, we calculated Kong and Cox [1997] lod scores at .5 cM intervals using Merlin [Abecasis et al., 2002]. We used the linear model [Kong and Cox, 1997] and the S_{Pairs} statistic [Whittemore and Halpern, 1994] for all analyses presented here; other non-parametric tests gave similar results. Over all replicates, we recorded the average lod score profile (which gives the ELOD at any analysis position), the empirical P-value for the lod score at the disease locus position, the average maximum lod score across each simulated chromosome (EMLOD), the average position of the maximum lod score (as an estimate of the disease locus location), and the proportion of maximum lod scores > 1 and > 3. In addition, we recorded the average width of the 1-lod support interval for the position of the disease locus and the proportion of those intervals that included the true disease locus position, d. We defined the width of the 1-lod support interval as $R - L$, where L and R are the first positions at which the lod score is less than the maximum lod score minus 1 to the left and right of the position of the maximum lod score, respectively. We recorded these measures for analyses that used (a) the sex-specific maps used to generate the data, (b) the sex-averaged map (the average of the true sex-specific maps), and (c) sex-averaged and sex-specific maps with error (see below).

Since the overall genetic length of the chromosome is not likely to be grossly under or

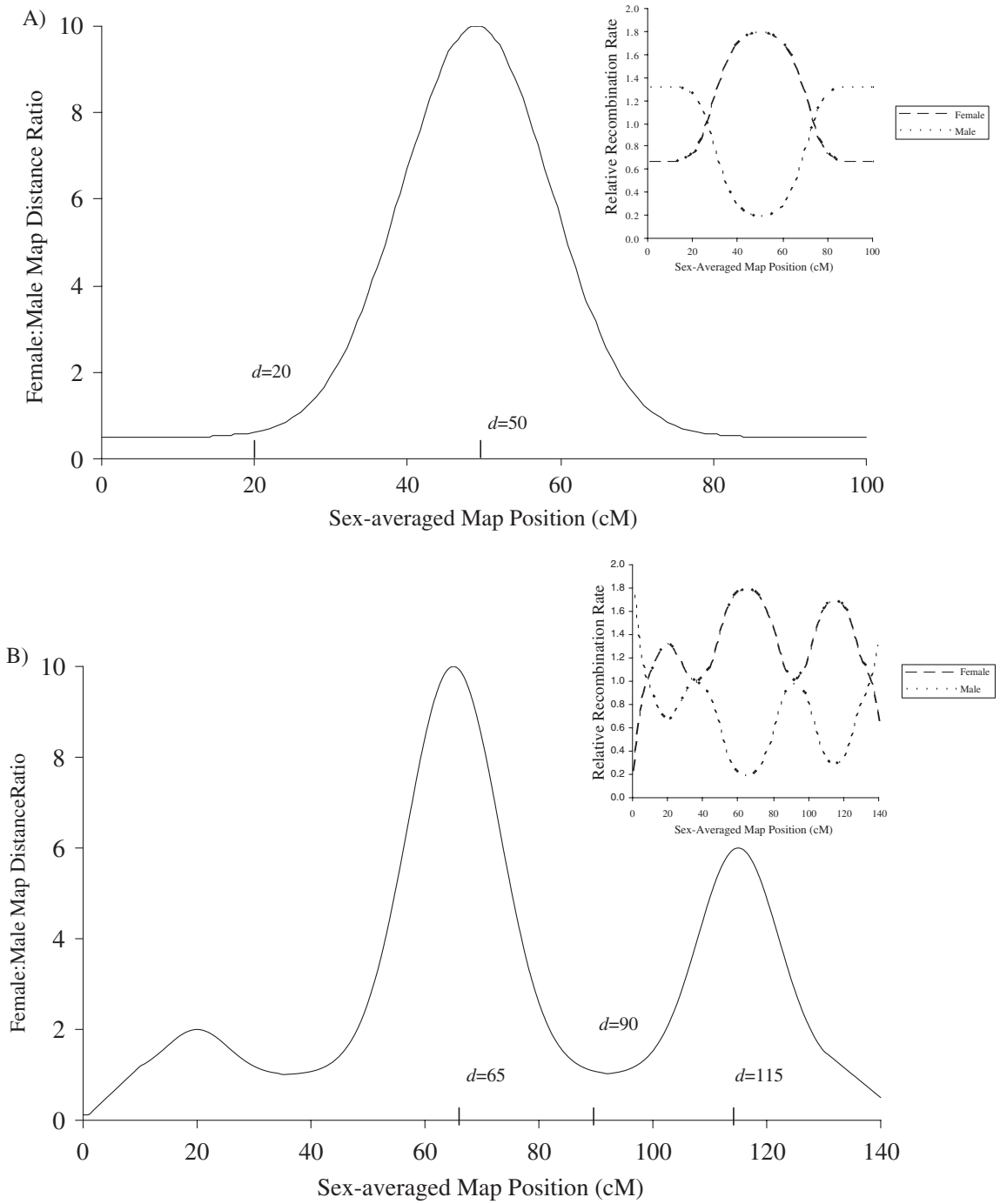

Fig. 1. Pattern of female:male map distance ratio as on (A) chromosome 19 and (B) chromosome 11 [Broman et al., 1998]. Position of disease locus, *d*, varied across simulations. Insets: Relative recombination rate for females and males for (A) chromosome 19 and (B) chromosome 11.

overestimated in practice, for (c) we fixed the chromosome length and varied the intermarker distances. We assumed that either 200 or 1,200 meioses were used to estimate the sex-averaged map positions, approximating the accuracy of the Marshfield [Broman et al., 1998] and deCODE [Kong et al., 2002] genetic maps, respectively. Note that since in our simulation all meioses were fully informative while the number of informative meiosis used to construct the Marshfield and deCODE maps varies along the genome, our simulated maps with error may be slightly more

accurate than the Marshfield and deCODE maps. We randomly distributed the expected number of recombination events for a given chromosome length and adjusted the map positions of each marker accordingly. For example, for a 100 cM map and 200 meioses, we expect 200 recombination events (2 per cM). We thus generated 200 uniformly distributed random numbers between 0 and 100 and counted the number of these which were $\leq m$, which we call a_m for each marker position m. Then a marker that was at position m cM was analyzed as if it mapped to $a_m/200$ cM. We carried out this process for each female:male ratio pattern to obtain a sex-averaged map with error. To generate the sex-specific maps with error, we used 2/3 of the random numbers for the female map and 1/3 for the male map since we assume that on average the number of recombination events for the female map is twice that for the male map [Broman et al., 1998]. We generated new analysis maps with error in this manner for each replicate data set.

To investigate the impact of missing parental genotype data, we performed the analysis four times for each replicate and analysis map, using the genotype data on 0, 1, or 2 parents. In the case that genotype data for only one parent were assumed missing, we initially excluded either all fathers or all mothers to examine the most extreme case of unequal representation of informative male and female meioses. To consider more common missing parental data patterns, we subsequently excluded only 25% or 50% of mothers or fathers. Finally, to investigate the impact of marker density, we also performed each analysis assuming markers were genotyped at 1 cM spacing, or instead, at 5 or 10 cM intervals.

IMPLEMENTATION OF MULTIPOINT LINKAGE ANALYSIS INCORPORATING SEX-SPECIFIC MAPS

The Lander and Green [1987] and Elston and Stewart [1971] algorithms can both incorporate sex-specific recombination fractions. While software packages implementing the Elston-Stewart algorithm or Monte-Carlo methods [Sobel and Lange, 1996; Heath, 1997] typically allow for sex-specific maps [e.g. Lathrop et al., 1984; O'Connell and Weeks, 1995; Schaffer, 1996], most packages implementing the Lander-Green algorithm do not. This is not due to a limitation of the algorithm, since good descriptions of how sex-specific recombination fractions can be incorporated in the analysis are available [Idury and Elston, 1997; Kruglyak and Lander, 1998], but merely for computational convenience.

We have modified the Merlin computer program [Abecasis et al., 2002] to allow for sex-specific recombination fractions in multipoint linkage analysis. Merlin uses the algorithm of Idury and Elston [1997] to carry out multipoint calculations and this enhancement mainly requires additional book keeping, so that two recombination fractions can be tracked for each interval. In addition, we implement code to automatically disable the founder-couple symmetry [Gudbjartsson et al., 2000] when sex-specific recombination fractions are in use. Overall, these changes do not result in a noticeable change in performance for most pedigrees, but do result in an increase in computing time and memory usage for pedigrees with a pair of ungenotyped and unphenotyped grandparents (since the number of inheritance vectors that must be enumerated in these pedigrees doubles when the founder-couple symmetry is disabled). Since these changes form part of the basic multipoint calculation, they naturally extend to all analyses supported by Merlin, including parametric and non-parametric linkage analysis, variance components and regression-based quantitative trait linkage analysis, haplotyping and genotype error detection.

RESULTS

The impact of using the sex-averaged map instead of sex-specific maps was very similar for the two patterns of r and the values of $\lambda_s > 1$ we considered. For brevity, unless otherwise noted, we report here the results for the chromosome 11 pattern (Fig. 1B) for $\lambda_s = 1$ and 1.25.

EFFECT OF USING SEX-AVERAGED MAP WHEN EITHER BOTH PARENTS OR NEITHER PARENT GENOTYPED

We first considered the situation where either all parental genotypes were available or all parental genotypes were missing. We found no meaningful differences in the average maximum LOD score (EMLOD, Table I) or average position of the MLOD between analyses that used the true sex-specific maps and those that used the sex-averaged map. For example, the EMLOD at 5-cM marker density was 2.59 when all parental genotypes were available for both maps when the ratio of female to male map distance $r = 10$

TABLE I. EMLOD under sex-averaged and sex-specific analysis maps for $\lambda_s = 1.25$

			Parents genotyped			
d	Marker density (cM)	Analysis map[a]	Both	Neither	Mother only	Father only
Female recombination rate higher than male recombination rate						
65	1	A	2.86	2.58	2.90	2.64
$r = 10$		S	2.86	2.70	2.81	2.81
	5	A	2.59	2.18	2.99	1.89
		S	2.59	2.26	2.45	2.46
	10	A	2.20	1.92	2.75	1.57
		S	2.22	1.89	2.06	2.09
90	1	A	2.85	2.68	2.82	2.77
$r = 1.1$		S	2.85	2.70	2.80	2.80
	5	A	2.59	2.21	2.71	2.20
		S	2.59	2.24	2.43	2.43
	10	A	2.40	1.99	2.69	1.81
		S	2.41	1.97	2.17	2.19
115	1	A	2.86	2.61	2.88	2.67
$r = 6$		S	2.86	2.69	2.80	2.80
	5	A	2.58	2.22	2.89	2.00
		S	2.59	2.26	2.45	2.45
	10	A	2.19	1.90	2.57	1.65
		S	2.20	1.87	2.05	2.05
Male recombination rate higher than female recombination rate[b]						
20	1	A	2.78	2.59	2.65	2.78
$r = .5$		S	2.78	2.62	2.72	2.73
	5	A	2.48	2.16	2.03	2.67
		S	2.48	2.19	2.35	2.37
	10	A	2.29	1.93	1.72	2.53
		S	2.29	1.92	2.09	2.10

[a]A = sex-averaged analysis map; S = sex-specific analysis map.
[b]Pattern (a) as for chromosome 19; see Figure 1A.

TABLE II. Empirical false-positive rates for $\lambda_s = 1.00$, $d = 65$ and $r = 10$

			Nominal significance level			
Parent genotyped	Marker density (cM)	Analysis map[a]	.05	.01	.001	.0001
Both	1	A	.056	.009	.0011	.0001
		S	.055	.009	.0011	.0001
	5	A	.057	.011	.001	.0001
		S	.057	.011	.001	.0001
	10	A	.055	.011	.0012	.0002
		S	.055	.011	.0012	.0001
Neither	1	A	.048	.008	.0003	.0002
		S	.058	.010	.0008	.0002
	5	A	.050	.009	.0005	.0001
		S	.057	.011	.0006	.0001
	10	A	.059	.010	.0007	$<.0001$
		S	.056	.009	.0008	$<.0001$
Mother	1	A	.064	.011	.0017	.0002
		S	.056	.009	.0013	.0002
	5	A	.118	.030	.0031	.0007
		S	.058	.010	.001	.0001
	10	A	.153	.043	.0057	.0006
		S	.055	.009	.0012	$<.0001$
Father	1	A	.043	.007	.0005	.0001
		S	.056	.009	.0011	.0002
	5	A	.017	.002	.0002	$<.0001$
		S	.059	.001	.0006	.0002
	10	A	.018	.002	$<.0001$	$<.0001$
		S	.057	.010	.007	$<.0001$

[a]A = sex-averaged analysis map; S = sex-specific analysis map.

(Table I, 4th column, 3rd and 4th data rows). The slight exception to this rule was observed only in the most extreme case of $r = 10$, where the sex-specific EMLOD is greater than that for the sex-averaged for 1-cM spacing when no parents are genotyped (Table I, 5th column, 1st and 2nd data rows). As expected based on the results for the EMLOD and disease locus position estimate, we also saw no meaningful differences between the analyses in the (1) false-positive rates for $\lambda_s = 1$ (Table II), (2) proportion of lod scores >1 or >3 (for replicates simulated under the alternative hypothesis where $\lambda_s > 1$, Table III), or (3) the proportion of 1-lod support intervals that included the true disease locus position d for $\lambda_s > 1$ (Table III).

The similar results for the two analysis maps in these two cases can be explained by the following. First, when both parents are genotyped, the map distances are less important in estimating IBD status and hence misspecifying the sex-specific distances by using the sex-averaged map is not detrimental. Second, when neither parent is genotyped, the sex-specific meioses cannot be identified, and the under-estimation

TABLE III. Percent of replicates with MLOD >1, MLOD >3, and with 1-lod support interval that includes the true disease locus position, d for $\lambda_s = 1.25$, $d = 65$ and $r = 10$

			Percent		
Parent genotyped	Marker density (cM)	Analysis map[a]	MLOD>1	MLOD>3	d in 1-lod support interval
Both	1	A	.95	.40	.72
		S	.95	.40	.72
	5	A	.92	.33	.86
		S	.92	.33	.86
	10	A	.85	.22	.88
		S	.85	.23	.87
Neither	1	A	.92	.32	.77
		S	.93	.36	.79
	5	A	.85	.23	.86
		S	.86	.25	.87
	10	A	.79	.16	.88
		S	.78	.15	.89
Mother	1	A	.95	.41	.75
		S	.94	.39	.75
	5	A	.95	.44	.87
		S	.89	.30	.86
	10	A	.93	.37	.89
		S	.83	.19	.88
Father	1	A	.93	.34	.74
		S	.94	.39	.75
	5	A	.79	.15	.83
		S	.90	.29	.86
	10	A	.68	.09	.86
		S	.83	.20	.88

[a]A = sex-averaged analysis map; S = sex-specific analysis map.

and over-estimation of the sex-specific map distances due to using the sex-averaged map approximately cancel out.

EFFECT OF USING SEX-AVERAGED MAP WHEN EITHER ALL MOTHERS OR ALL FATHERS GENOTYPED

In contrast to the case of having either all parental genotypes or none, we did see differences in the EMLOD between analyses using sex-specific maps and the sex-averaged map when only paternal or only maternal genotypes were available.

We found that for all of the cases with a higher female recombination rate ($r > 1$) we considered, the ELODs and EMLOD were inflated when the sex-averaged map was used for analysis but only mothers were genotyped and the marker density was low (5 and 10 cM). Conversely, the ELODs and EMLOD were deflated when the sex-averaged map was used for analysis but only fathers were genotyped. Differences in the ELODs and EMLOD based on analysis maps with 1-cM marker densities were generally very small, probably because IBD is essentially known in this case. Fig. 2 shows the ELOD for each analysis position over all replicates under the null hypothesis of $\lambda_s = 1$ across the different marker densities when either all maternal or all paternal genotypes are missing. The inflation or deflation in the ELOD for the 5 and 10 cM cases mirrors the pattern of the ratio of male and female map distance (r) and is greatest where r is farthest from 1. The inflation of the ELOD when sex-averaged maps were used for analysis and female recombination rates were higher ($r > 1$) but only maternal genotypes were available led to increased false-positive rates at 5 and 10 cM density (Table II). In the same setting, but when only paternal genotypes were available, false-positive rates decreased (Table II) and ELODs were deflated.

Similar to Fig. 2, Fig. 3 shows the ELOD for each analysis position over all replicates under the alternative hypothesis of $\lambda_s = 1.25$ for $d = 65$ across the different marker densities when either all maternal or all paternal genotypes are missing. The inflation or deflation of the ELOD for the 5 and 10 cM cases is greatest at the position of the disease locus, which in this case corresponds to the maximum of r. Table I shows the mean maximum LOD score (EMLOD) when $\lambda_s = 1.25$ for each of the three alternative disease-locus positions we simulated. For example, when the ratio of female to male map distance $r = 10$, the EMLOD using sex-averaged maps when only mothers are genotyped is 2.99 at 5-cM marker density while the EMLOD using sex-specific maps is only 2.45 (Table I, 6th column, 3rd and 4th rows). Notably, even for a modest $r = 1.1$ ($d = 90$), changes in the EMLOD due to the sex-averaged map are evident. When sex-specific maps were assumed, the EMLODs for the analyses incorporating only maternal or only paternal genotypes were nearly identical and intermediate to those for the settings where all parents were genotyped or all parents were

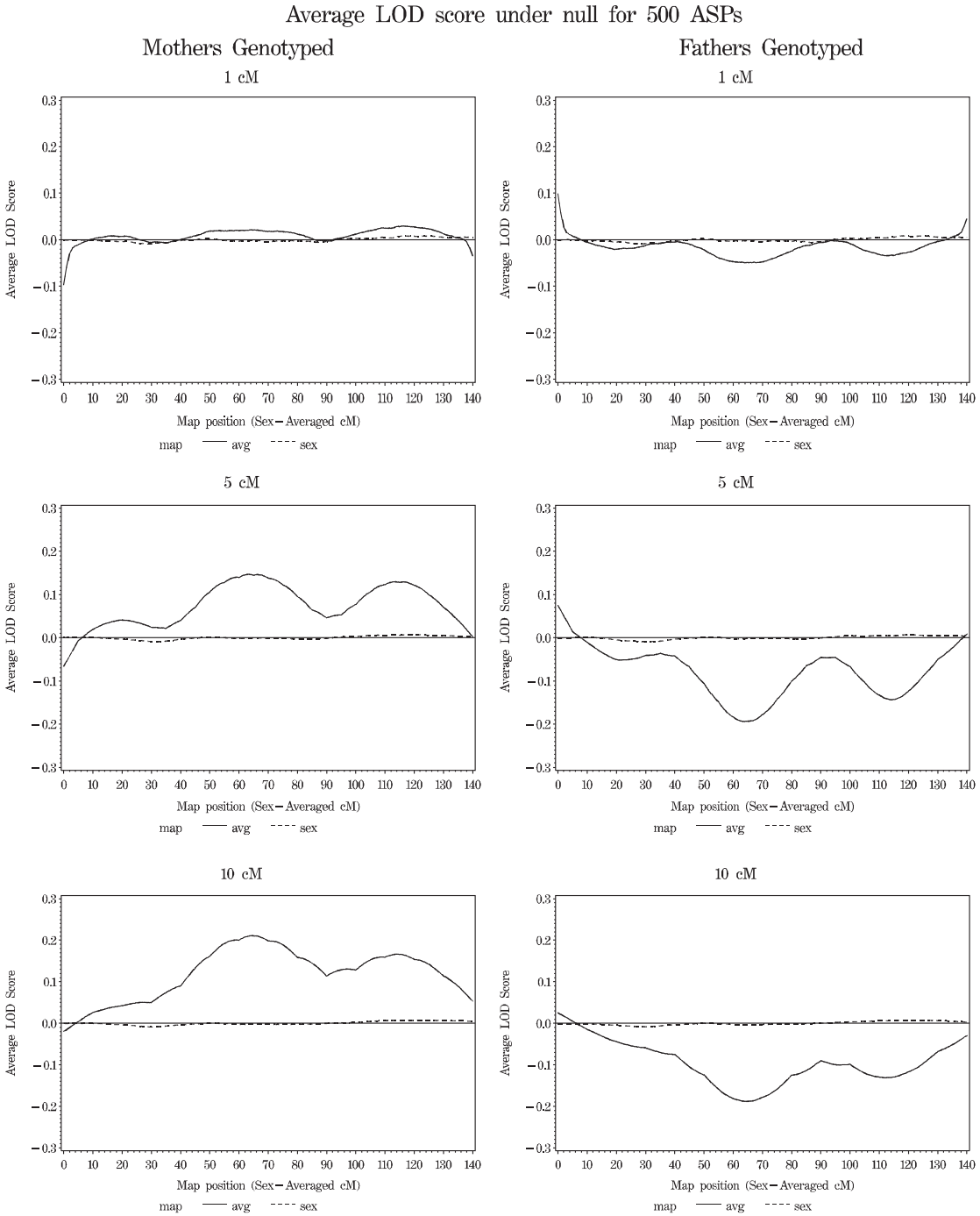

Fig. 2. Average LOD score over all replicates (ELOD) under the null hypothesis ($\lambda_s = 1.00$) for decreasing marker density when either all mothers or all fathers genotyped.

missing. Taken together with the false-positive rate results, these results indicate that the EMLODs using the sex-specific map were of the appropriate magnitude and that the EMLODs using the sex-averaged maps were inappropriately inflated when only mothers were genotyped and deflated when only fathers were genotyped.

The differences in the ELOD and EMLOD at the 5 and 10 cM marker densities can be explained by the fact that the sex-averaged map underestimates the female map distances and overestimates the male map distances for $r > 1$. When all mothers are genotyped, the female meioses are well inferred and the sex-averaged map distances are primarily

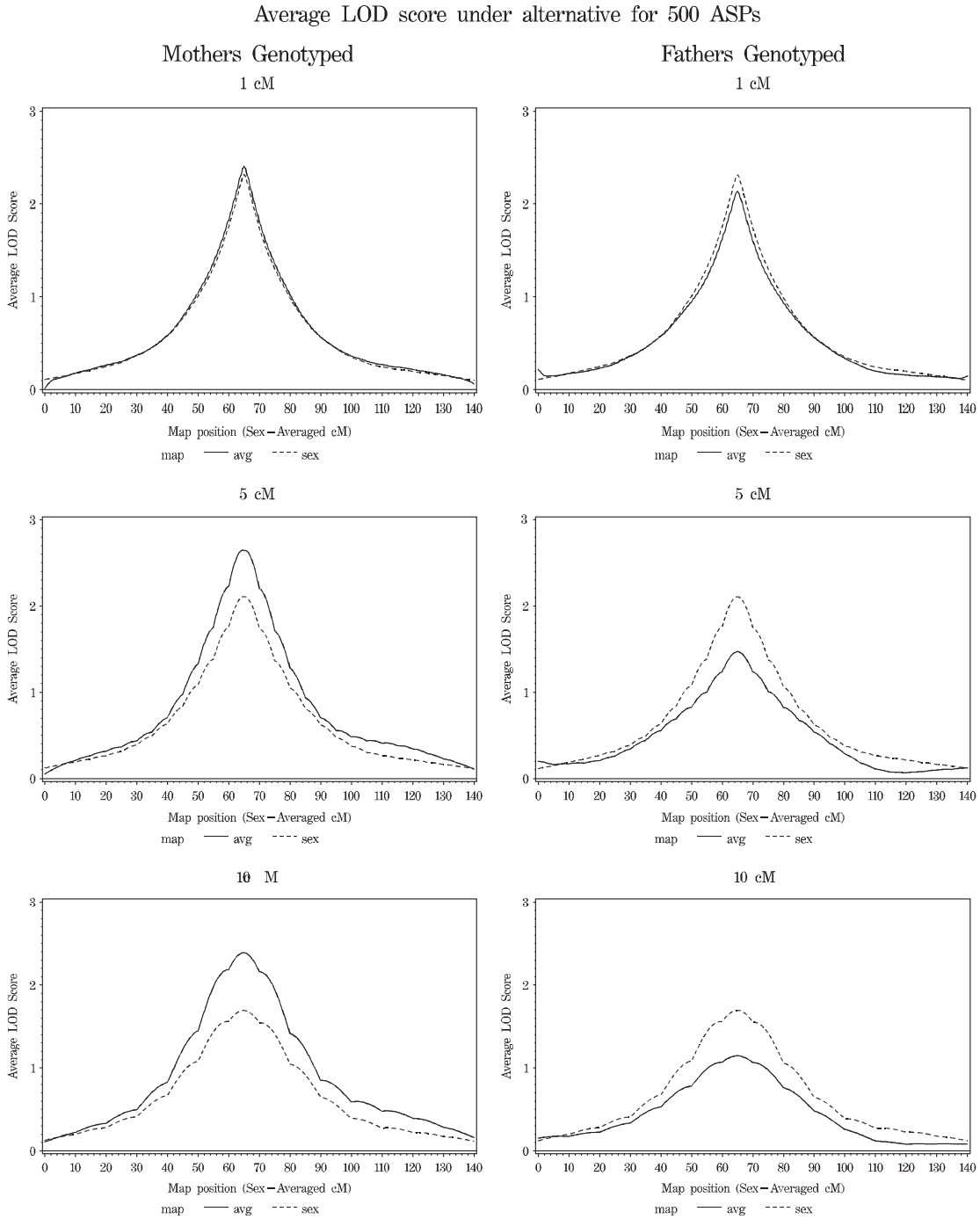

Fig. 3. Average LOD score over all replicates (ELOD) under the alternative hypothesis ($\lambda_s = 1.25$) for decreasing marker density when either all mothers or all fathers genotyped.

influential in estimating the IBD states for male meioses. Since the sex-averaged map overestimates the male recombination fractions, more recombination events are modeled than actually occur and this results in an inflated EMLOD (because IBD is estimated to increase too rapidly flanking each locus where IBD = 0 is observed). The converse holds when all fathers are genotyped since underestimating the recombination rate for female meioses leads to inferring too few recombination events and therefore inferring that large swathes of the chromosome

TABLE IV. EMLOD under sex-averaged and sex-specific analysis maps for $\lambda_s = 1.25$, $r = 10$ under less extreme maternal genotype imbalance for sib pairs

Marker density (cM)	Analysis map[a]	% Mothers gentoyped 0%	25%	50%	100%
1	A	2.58	2.66	2.74	2.90
	S	2.70	2.73	2.76	2.81
5	A	2.18	2.37	2.57	2.99
	S	2.26	2.31	2.35	2.45
10	A	1.92	2.12	2.32	2.75
	S	1.89	1.94	1.98	2.06

[a]A = sex-averaged analysis map; S = sex-specific analysis map.
Note: No (0%) fathers genotyped.

are IBD = 0. As expected, for $r < 1$ as in pattern (a), $d = 20$ case ($r = .5$), we saw the opposite trends (Table I, last six rows).

Despite the inflation and deflation in the EMLOD due to using the sex-averaged map, we did not observe large differences in the accuracy or precision of localization of the disease locus when sex-averaged maps were used instead of sex-specific maps (data not shown). In addition, when the imbalance in parental genotypes was reduced so that 25% or 50% of mothers were genotyped and no fathers were genotyped and vice versa, the changes in the EMLOD (and hence the false-positive rate) due to the sex-averaged analysis map were attenuated in an approximately linear manner (Table IV). The percentage increase in the EMLOD when only mothers are genotyped was 33% for $\lambda_s = 1.25$ and $r = 10$ at 10 cM marker density, but this was reduced to 17% and 9% when 50% and 25% of mothers (and no fathers) were genotyped, respectively.

EFFECT OF ANALYSIS MAPS WITH STOCHASTIC ERROR

Simulations using maps estimated from 200 or 1200 fully informative meiosis resulted in nearly identical conclusions to the simulations described above, which benefit from the sex-specific and sex-averaged maps used to generate the simulated data. As expected, when any parental genotypes were missing, the magnitude of the EMLODs was reduced due to map inaccuracies for maps estimated from 200 fully informative meioses. However, the reduction was similar for the sex-specific and sex-averaged map analyses. Specifically, we observed no situations where using the sex-specific maps resulted in a loss of power, even though only 100 meioses are available to estimate the male and female genetic maps in this setting. When only mothers were genotyped, the percentage inflation in the EMLOD resulting from sex-average maps was 38% for $\lambda_s = 1.25$ and $r = 10$ at 10 cM marker density and assuming 200 meiosis are used to estimate the genetic map. For maps estimated from 1200 fully informative meioses, we saw no reduction in the EMLODs due to map inaccuracies.

EFFECT OF SNP MARKERS WITH ALLELE FREQUENCY = .50

Results for the 5- and 10-cM SNP density were qualitatively very similar to the 5- and 10- cM microsatellite marker density, although the ELOD was inflated by only 16% when only mothers were genotyped for $\lambda_s = 1.25$ and $r = 10$ at 10 cM marker density. The negative impact of using the sex-averaged map was evident even at the 1 cM SNP density, likely reflecting the reduction in IBD information provided by the SNP markers at this density.

EFFECT OF SIBSHIP SIZE

We saw a very similar pattern of results for sibship sizes 3, 4, and 5 as we saw for sib pairs. As the sibship size increased, however, the inflation (deflation) of the EMLOD decreased (Table V), presumably as a result of the increase in inheritance information obtained with the additional siblings. The percentage increase in the EMLOD when only mothers are genotyped was 24%, 15%, and 9% for affected sibship sizes of 3, 4 and 5, respectively, for $\lambda_s = 1.25$ and $r = 10$ at 10 cM marker density.

EFFECT OF SAMPLE SIZE

Differences in the number of sib pairs did not alter the pattern of effects due to missing parental genotypes, although the percentage increase in the EMLOD was slightly increased for the sample size of 1000 and slightly decreased for the sample size of 250 compared to 500. The percentage increase in the EMLOD when only mothers are genotyped was 25% and 39% for samples of 250 and 1000 ASPs, respectively, for $\lambda_s = 1.25$ and $r = 10$ at 10 cM marker density.

TABLE V. EMLOD under sex-averaged and sex-specific analysis maps for $\lambda_s = 1.25$, $d = 65$ and $r = 10$ for larger sibships

Sibs	Marker density (cM)	Analysis map[a]	Parents genotyped			
			Both	Neither	Mother	Father
3	1	A	4.17	3.86	4.20	3.93
		S	4.17	4.01	4.13	4.10
	5	A	3.86	3.16	4.22	2.90
		S	3.87	3.29	3.69	3.58
	10	A	3.23	2.65	3.73	2.34
		S	3.26	2.62	3.02	2.97
4	1	A	4.61	4.42	4.62	4.46
		S	4.61	4.50	4.58	4.55
	5	A	4.28	3.60	4.48	3.48
		S	4.29	3.72	4.14	3.97
	10	A	3.59	2.86	3.85	2.76
		S	3.61	2.88	3.36	3.26
5	1	A	4.40	4.31	4.41	4.31
		S	4.40	4.34	4.39	4.36
	5	A	4.08	3.59	4.20	3.54
		S	4.09	3.68	4.00	3.84
	10	A	3.46	2.78	3.57	2.85
		S	3.48	2.86	3.29	3.18

[a]A = sex-averaged analysis map; S = sex-specific analysis map.

EFFECT OF PEDIGREE STRUCTURE: COUSIN-PAIR AND HALF-SIB FAMILIES

The effects of missing parental genotypes for affected cousin pairs were qualitatively similar to those for full sibs whether the cousins were offspring of (1) a pair of sisters, (2) a pair of brothers, or (3) a brother and a sister (data not shown). We examined affected half-sib pairs which shared their mother or which shared their father. When the genotypes for the parent in common were available, the sex-specific and sex-averaged EMLODs were very similar. This was expected since information on IBD status is only gained from the shared parent (who in this case is genotyped), and hence the sex-averaged map is not influential for IBD estimation. When genotypes for the shared parent were missing (either because all parental genotypes were missing or because genotypes are available only for the parent not in common between the half-sibs), we saw the same patterns as described for full sibs when genotypes for one parent were missing (data not shown).

DISCUSSION

Large differences in male and female recombination rates across the human genome are well documented [Straub et al., 1993; Broman et al., 1998; Mohrenweiser et al., 1998; Kong et al., 2002], but the biological mechanism(s) for these differences are not yet well understood. Since maternally and paternally inherited chromosomes have the same ordering of genes and intervening sequence, sequence characteristics cannot explain sex differences in recombination rates. The mechanisms of genomic imprinting, including differential methylation and/or heterochromatin structure formation [Pfeifer, 2000], are associated with sex-specific differences in recombination rate [Paldi et al., 1995]. Perhaps DNA methylation patterns and heterochromatin formation differ between chromosomes in oocytes and spermatocytes so that different regions of a chromosome are available for chiasma formation in each case [Robinson, 1996]. Another hypothesis is that genes which regulate crossing over [Robinson, 1996] are differentially regulated in males and females.

Even though we do not understand the reasons for sex-specific variation in recombination, it has been a cause for concern for those attempting to map and identify complex disease susceptibility loci using linkage analysis. Underestimation of the intermarker distances in a multipoint setting can decrease power to detect linkage [Halpern and Whittemore, 1999], while over-estimation can increase the false-positive rate when parental genotypes are missing [Hauser, 1998]. Since a sex-averaged map generally underestimates one sex-specific map and overestimates the other, using a sex-averaged map in place of sex-specific maps has the potential to compromise the testing properties of linkage analysis.

We chose two representative patterns of variation in female:male recombination rates. Pattern (a) represents the most common pattern with a peak in relative female recombination rate around the centromere and a dip near the telomeres; pattern (b) represents the case of a more variable ratio along the chromosome with two distinct peaks in the relative recombination rate in females that may influence disease locus localization. We found that when either all or no parental genotypes were available, the EMLOD and esti-

mates of the disease locus position were virtually identical irrespective of whether sex-specific maps or a sex-averaged map were used for analysis. In contrast, when there was an imbalance in the number of mothers and fathers genotyped, the size of the test for linkage could be inflated or power deflated depending on whether more mothers or fathers were genotyped and whether the female:male genetic map-distance ratio was greater than or less than 1. The imbalance in parental genotypes had very similar effects in larger sibship sizes and affected cousin-pair families.

While the negative implications of using a sex-averaged map were reduced when the imbalance was less severe (e.g. when no fathers, but only 50% of mothers were genotyped), these results suggest that linkage analyses with sex-specific maps should be performed when possible if any imbalance in parental genotypes exists. For linkage studies of adult probands, such imbalances are common, both due to differences in survival rates between the sexes and also due to differences in participation rates. For example, in the BetaGene project [Watanabe et al., 2005], a linkage study of relatives of women with gestational diabetes, the ratio of enrolled mothers to fathers is 108:54, despite attempts to recruit both mothers and fathers [Thomas Buchanan, personal communication]. These imbalances may be even larger in epidemiological studies of diseases affecting primarily one sex, such as breast and prostate cancer.

Our simulations did not suggest that the increased sampling error associated with the sex-specific maps resulted in a loss of efficiency. If the number of meioses used to create the genetic maps for a given analysis is very limited, and increased sampling error is of particular concern, then a sex-specific map analysis can serve as a sensitivity analysis. If there is a large discrepancy between the sex-averaged map and sex-specific map results, then further work can be done to examine the female:male map distance ratio in the appropriate region of the genome using published maps [e.g. http://research.marshfieldclinic.org/genetics/Map_Markers/maps/IndexMapFrames.html, Kong et al., 2002; Kong et al., 2004]. Since ASPEX [Hinds and Risch, 1999], Allegro [Gudbjartsson et al., 2000] and Merlin [Abecasis et al., 2002] readily incorporate sex-specific maps for analysis, this strategy provides a relatively simple way to avoid any increase in the false-positive rate or loss of power due to differences in the recombination rate between males and females.

We have not examined the special case of imprinting. Recently, there has been renewed interest in testing for parent-of-origin effects as part of a linkage analysis [Paterson et al., 1999; Strauch et al., 2000; Hanson et al., 2001; Shete and Amos, 2002]. Lindsay et al. [2001] recently presented an application of parent-of-origin quantitative linkage analysis that used sex-averaged maps. While the authors described limited simulations to evaluate the impact of ignoring a 5:1 female:male map distance ratio for their data, a thorough investigation of the impact of using sex-averaged maps has not been conducted. Because these methods test specifically for a parent-of-origin effect, and sex-specific differences in recombination rate can mimic parent-of-origin effects [Paldi et al., 1995], it seems that the potential for elevated type I error rates due to using a sex-averaged map is enhanced in this setting.

In conclusion, we have compared the EMLOD and power and size of tests for linkage to a dichotomous trait locus when the correct sex-specific maps are used to these quantities when the sex-averaged map is used for analysis in the presence of differences between the male and female recombination rates. We found no situations where it was harmful to use the sex-specific maps, even after taking into account the higher uncertainty that goes into the estimation of each recombination fraction for these maps. However, we did find situations where using the sex-averaged map resulted in inflated false-positive rates or reduced power. Thus, we recommend that sex-specific maps be used for linkage analysis whenever possible.

ACKNOWLEDGMENTS

This research was supported by National Institutes of Health Grants HG00376 (to M.B.) and HG02651 (to G.R.A.). T.E.F. was previously supported by National Institutes of Health Training Grant HG00040.

REFERENCES

Abecasis GR, Cherny SS, Cookson WO, Cardon LR. 2002. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30:97–101.

Broman KW, Murray JC, Sheffield VC, White RL, Weber JL. 1998. Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am J Hum Genet 63: 861–869.

Daw EW, Thompson EA, Wijsman EM. 2000. Bias in multipoint linkage analysis arising from map misspecification. Genet Epidemiol 19:366–380.

Elston RC, Stewart J. 1971. A general model for the genetic analysis of pedigree data. Hum Hered 21:523–542.

Gudbjartsson DF, Jonasson K, Frigge ML, Kong A. 2000. Allegro, a new computer program for multipoint linkage analysis. Nat Genet 25:12–13.

Haldane JBS. 1919. The combination of linkage values and the calculation of distances between the loci of linked factors. J Genet 8:299–309.

Halpern J, Whittemore AS. 1999. Multipoint linkage analysis. A cautionary note. Hum Hered 49:194–196.

Hanson RL, Kobes S, Lindsay RS, Knowler WC. 2001. Assessment of parent-of-origin effects in linkage analysis of quantitative traits. Am J Hum Genet 68:951–962.

Hauser ER. 1998. Methods for linkage analysis of complex genetic disease. Ph.D. dissertation, University of Michigan.

Heath SC. 1997. Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. Am J Hum Genet 61:748–760.

Hinds D, Risch N. 1999. The ASPEX package: affected sib-pair mapping. Unpublished computer documentation.

Idury RM, Elston RC. 1997. A faster and more general hidden Markov model algorithm for multipoint likelihood calculations. Hum Hered 47:197–202.

Kong A, Cox NJ. 1997. Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 61:1179–1188.

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K. 2002. A high-resolution recombination map of the human genome. Nat Genet 31:241–247.

Kong X, Murphy K, Raj T, He C, White PS, Matise TC. 2004. A combined linkage-physical map of the human genome. Am J Hum Genet 75:1143–1148.

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. 1996. Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363.

Kruglyak L, Lander ES. 1998. Faster multipoint linkage analysis using Fourier transforms. J Comput Biol 5:1–7.

Lander ES, Green P. 1987. Construction of multilocus genetic linkage maps in humans. Proc Natl Acad Sci USA 84:2363–2367.

Lathrop GM, Lalouel JM, Julier C, Ott J. 1984. Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81:3443–3446.

Lindsay RS, Kobes S, Knowler WC, Bennett PH, Hanson RL. 2001. Genome-wide linkage analysis assessing parent-of-origin effects in the inheritance of type 2 diabetes and BMI in Pima Indians. Diabetes 50:2850–2857.

Mohrenweiser HW, Tsujimoto S, Gordon L, Olsen AS. 1998. Regions of sex-specific hypo- and hyper-recombination identified through integration of 180 genetic markers into the metric physical map of human chromosome 19. Genomics 47: 153–162.

NIH/CEPH Collaborative Mapping Group. 1992. A comprehensive genetic linkage map of the human genome. Science 258: 148–162.

O'Connell JR, Weeks DE. 1995. The VITESSE algorithm for rapid exact multilocus linkage analysis via genotype set-recoding and fuzzy inheritance. Nat Genet 11:402–408.

Paldi A, Gyapay G, Jami J. 1995. Imprinted chromosomal regions of the human genome display sex-specific meiotic recombination frequencies. Curr Biol 5:1030–1035.

Paterson AD, Naimark DM, Petronis A. 1999. The analysis of parental origin of alleles may detect susceptibility loci for complex disorders. Hum Hered 49:197–204.

Pfeifer K. 2000. Mechanisms of genomic imprinting. Am J Hum Genet 67:777–787.

Robinson WP. 1996. The extent, mechanism, and consequences of genetic variation, for recombination rate. Am J Hum Genet 59: 1175–1183.

Schaffer AA. 1996. Faster linkage analysis computations for pedigrees with loops or unused alleles. Hum Hered 46: 226–235.

Shete S, Amos CI. 2002. Testing for genetic linkage in families by a variance-components approach in the presence of genomic imprinting. Am J Hum Genet 70:751–757.

Sobel E, Lange K. 1996. Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. Am J Hum Genet 58:1323–1337.

Straub RE, Speer MC, Luo Y, Rojas K, Overhauser J, Ott J, Gilliam TC. 1993. A microsatellite genetic linkage map of human chromosome 18. Genomics 15:48–56.

Strauch K, Fimmers R, Kurz T, Deichmann KA, Wienker TF, Baur MP. 2000. Parametric and nonparametric multipoint linkage analysis with imprinting and two-locus-trait models: application to mite sensitization. Am J Hum Genet 66: 1945–1957.

Watanabe RM, Xiang AH, Allayee H, Hartiala J, Trigo E, Wang C, Berrios F, Hernandez, J, Paredes G, Hernandez M, Cercado S, Patel L, Caro J, Kjos SA, Lawrence JM, Buchanan TA. 2005. Variation in the P2-promoter region of hepatocyte nuclear factor-4A (***HNF4A***) is associated with β-cell function in Mexican American (MA) families of a proband with gestational diabetes (GDM). Am Diab Assoc Annual Meeting Abstract 153.

Whittemore AS, Halpern J. 1994. A class of tests for linkage using affected pedigree members. Biometrics 50:118–127.