Table 1. Type I Error Rates for Multiplex Families

Method[b]	Family Types[a]	
	Three Children	Four Children
X-APL	0.056	0.052
XMCPDT$_T$	0.049	0.042
XMCPDT$_E$	0.049	0.041

[a] Two multiplex family scenarios were considered. Three and four children refer to families with three and four children, respectively, with at least two of them being affected.
[b] XMCPDT$_T$ and XMCPDT$_E$ refer to XMCPDT with true and estimated allele frequencies, respectively.

addition to the simulation detailed above, we also considered a hypothetical study focusing on multiplex families. We considered two scenarios, one with three children, two or three of them being affected, and the other with four children, at least two of them being affected. Either scenario clearly violated the sampling assumption, but the violation was not severe because not all families were forced to have exactly the same number of affected children. For both the three-children and the four-children families, XMCPDT with either true or estimated allele frequencies gave a p value of less than 0.05 (the nominal), demonstrating once again its robustness to slight departure from the assumption (Table 1). These results were based on 100 simulated families with the RecA model[1] and 4000 replicated runs. For each run, half of the families were assumed to have missing parental genotypes. We chose to perform much longer runs to obtain more accurate estimates of the actual type I error rates.

In contrast, for datasets with extended pedigrees, X-APL tends to have inflated type I error rates. The reason might be that when handling extended pedigrees, X-APL dissects them into nuclear families and analyzes them as if they were independent. However, whether this is the main reason remains unclear because explicit explanation on how extended pedigrees were handled was not available in Chung et al.[1] It is clear, though, that X-APL is a valid test only for nuclear families, and as such, it should not come as a surprise that it has inflated type I error rates when used for analysis of data from extended pedigrees. Perhaps X-APL and XMCPDT should not be viewed as competing approaches; rather, they should be viewed as complementary, utilizing their individual strengths. In particular, X-APL could be used for analyzing data from nuclear families, whereas data from extended pedigrees might be better treated with XMCPDT. For a dataset comprising both types of family, a combined analysis utilizing the strengths of both methods would be desirable.

Jie Ding[1] and Shili Lin[1,*]
[1]Department of Statistics, Ohio State University, Columbus, OH 43210, USA
*Correspondence: shili@stat.osu.edu

Acknowledgments

This work was supported in part by National Institutes of Health grant HG002657 and the Ohio State BRTT grant Tech 05-062.

Web Resources

The URLs for data presented herein are as follows:

X-APL, http://www.chg.duke.edu/research/software.html
XMCPDT, http://www.stat.osu.edu/~statgen/SOFTWARE/MC-PDT/

References

1. Chung, R., Morris, R.W., Zhang, L., Li, Y., and Martin, E.R. (2007). X-apl: An improved family-based test of association in the presence of linkage for the X chromosome. Am. J. Hum. Genet. *80*, 59–68.
2. Ding, J., Lin, S., and Liu, Y. (2006). Monte carlo pedigree disequilibrium test for markers on X chromosome. Am. J. Hum. Genet. *79*, 567–573.

DOI 10.1016/j.ajhg.2007.11.011.

Response to Ding and Lin

To the Editor: In Chung et al.,[1] we reported simulation results showing that when a large fraction of families are missing parental genotypes, XMCPDT[2] can exhibit an inflated type I error rate. Ding and Lin dismiss the fraction of missing parental genotypes as an explanation for excess type I error and instead attribute our observation to violation of a sampling assumption of XMCPDT. They point out that our simulations condition on a fixed number of affected and unaffected offspring and note that this violates the XMCPDT assumption that family structure is random with respect to the number of affected offspring. To investigate this further, we performed a simulation study that allowed a variety of nuclear-family structures and varied the proportion of missing parent genotypes. Replicates of 300 families, each with three siblings, were generated via SIMLA[3] under an X-linked recessive disease model (RecF[1]). To ensure a variety of family phenotypes, we set disease prevalence to 0.3 and randomly sampled families with at least one affected sibling. Among 3000 replicates, the average proportions of families with one affected and two unaffected siblings, two affected and one unaffected siblings, and three affected siblings were 48%, 42%, and 10%, respectively. We believe that this simulation model achieves the family-ascertainment assumption of Ding et al.[2]

Figure 1 plots the relationship between type I error rate and the fraction of missing parental genotypes for XMCPDT, XPDT, and X-APL. Type I error rate increases

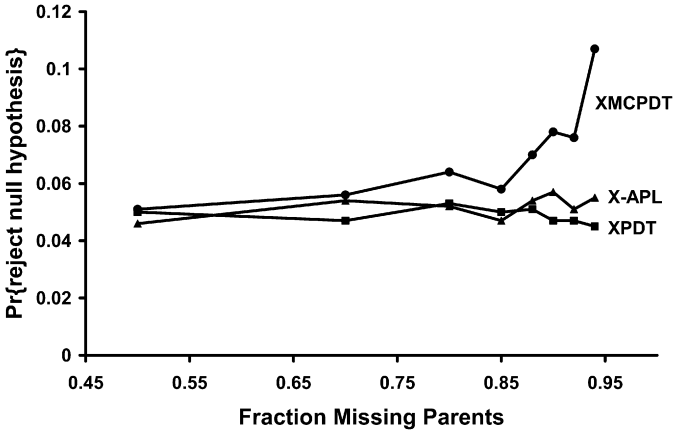

Figure 1. Type I Error Rates for XMCPDT, XPDT, and X-APL with Different Proportions of Families with Missing Parent Genotypes

In each sample, we generated different proportions of missing parents (94%, 92%, 90%, 88%, 85%, 80%, 70%, and 50%) by randomly removing parental genotypes regardless of family structure. Marker-allele frequency was 0.25. The marker and disease loci were tightly linked and in linkage equilibrium.

with the fraction of missing parents for XMCPDT but remains near the nominal level of 0.05 for XPDT and X-APL. These simulation results show that even in the presence of a variety of family types, the type I error rate of XMCPDT can be inflated when there is a large proportion of families with missing parents.

Ding and Lin present simulation results showing that the type I error rate for X-APL can be inflated for some models. These results are based on small sample sizes: 100 nuclear families in one simulation and 81 pedigrees in a simulation based on Ohio State University multiple sclerosis (OSUMS) pedigree structures. We have noted elsewhere[4] that APL, a test for autosomal markers that X-APL extends to sex-linked markers, can have an inflated type I error rate as a result of deviation of the test statistic from the standard normal distribution in small samples or in the presence of low marker-allele frequencies. Hence, it is not surprising that the small sample sizes used by Ding and Lin produced inflated type I error rates for X-APL. Moreover, APL is more sensitive to the small sample sizes than PDT (data not shown). To investigate the effect of small sample size on X-APL, we simulated 100 and 60 nuclear families with three siblings, where families were ascertained with at least one affected sibling and all parent genotypes were missing. Table 1 shows that type I error rates for X-APL can be slightly inflated in samples of 100 or 60 nuclear families. For the APL statistic, Chung et al.[4] noted that when the bootstrap variance exceeds 5, the level of the test based on the asymptotic distribution generally holds. Simulation results in Table 1 illustrate that this guideline can protect X-APL from being liberal in small samples.

Bearing in mind limitations of sample size and the extent of missing parental genotypes, we recommend, along with Ding and Lin that X-APL be used for samples consisting of nuclear families and that XMCPDT or XPDT be used for samples consisting of extended pedigrees. Moreover, we endorse their statement that an analysis combining the strengths of both methods would be desirable.

Table 1. Type I Error Rates for X-APL in Small Samples

Model	Type I Error	Adjusted Type I Error[a]
100 A++[b]		
RecA[c]	0.056	0.048 (88%[d])
RecB	0.058	0.050 (88%)
MultA	0.051	0.041 (88%)
MultB	0.055	0.048 (88%)
60 A++		
RecA	0.057	0.031 (8.5%)
RecB	0.058	0.030 (8%)
MultA	0.058	0.036 (9.1%)
MultB	0.059	0.030 (8.4%)

Type I error rates based on 10,000 replicates for 100 A++ and 50,000 replicates for 60 A++. Marker-allele frequency was 0.25.

[a] Type I error is the proportion of data sets with p value < 0.05 where only statistics with variance > 5 were evaluated.

[b] Families were ascertained with at least one affected sibling, and each family has three siblings.

[c] Disease models defined in Chung et al.[1]

[d] Proportion of data sets with variance for the X-APL statistic > 5.

Ren-Hua Chung,[1] Richard W. Morris,[2] and Eden R. Martin[2,*]

[1]Center for Human Genetics, Duke University Medical Center, Durham, NC 27710, USA; [2]Miami Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL 33136, USA

*Correspondence: emartin1@med.miami.edu

References

1. Chung, R.H., Morris, R.W., Zhang, L., Li, Y.J., and Martin, E.R. (2007). X-APL: An improved family-based test of association in the presence of linkage for the X chromosome. Am. J. Hum. Genet. *80*, 59–68.
2. Ding, J., Lin, S.L., and Liu, Y. (2006). Monte Carlo pedigree disequilibrium test for markers on the X chromosome. Am. J. Hum. Genet. *79*, 567–573.
3. Schmidt, M., Hauser, E.R., Martin, E.R., and Schmidt, S. (2005). Extension of the SIMLA package for generating pedigrees with complex inheritance patterns: Environmental covariates, gene-gene and gene-environment interaction. Stat. Appl. Genet. Mol. Biol. *4*, Article 15.
4. Chung, R.H., Hauser, E.R., and Martin, E.R. (2006). The APL Test: Extension to General Nuclear Families and Haplotypes and Examination of Its Robustness. Hum. Hered. *61*, 189–199.

DOI 10.1016/j.ajhg.2008.01.001.