Table 1. Performance of log-linear model under Carayol et al's (2006) simulation scenarios (Table I of Carayol et al 2006) using 500 replicates of 1000 triads

Carayol's Model	Number of SNPs in haplotype	Risk haplotype frequency	R_1 used for generating simulated data	R_2 used for generating simulated data	Pattern of missingness*	Estimated risk haplotype frequency	Geometric mean of estimated R_1 (95% CI)	Geometric mean of estimated R_2 (95% CI)	Empirical coverage of nominal 95% confidence interval for R1	Empirical coverage of nominal 95% confidence interval for R2
1	3	0.4	1.5	4	No missing	0.40	1.51 (1.5 ,1.52)	4.02 (3.98 ,4.07)	0.94	0.944
					5.3% missing	0.40	1.51 (1.49 ,1.52)	4.01 (3.97 ,4.06)	0.954	0.944
					15% missing	0.40	1.51 (1.5 ,1.52)	4.03 (3.98 ,4.08)	0.94	0.93
2	3	0.1	4	11	No missing	0.10	4.03 (3.99 ,4.06)	11.14 (10.94 ,11.34)	0.946	0.94
					5.3% missing	0.10	4.03 (4 ,4.07)	11.15 (10.95 ,11.36)	0.956	0.942
					15% missing	0.10	4.04 (4 ,4.07)	11.15 (10.93 ,11.37)	0.94	0.93
3	6	0.4	1.5	4	No missing	0.40	1.5 (1.49 ,1.51)	4.01 (3.97 ,4.06)	0.968	0.954
					2.7% missing	0.40	1.5 (1.49 ,1.51)	4.01 (3.97 ,4.06)	0.962	0.958
					15% missing	0.40	1.5 (1.49 ,1.51)	4.02 (3.97 ,4.06)	0.97	0.946
4	6	0.1	4	11	No missing	0.10	4 (3.96 ,4.04)	10.9 (10.72 ,11.09)	0.936	0.944
					2.7% missing	0.10	3.99 (3.96 ,4.03)	10.87 (10.69 ,11.07)	0.942	0.94
					15% missing	0.10	4.01 (3.97 ,4.05)	10.96 (10.76 ,11.16)	0.936	0.948

*This column represents the percentage of genotypes missing randomly, which is different from the missingness defined by Carayol et al. (the probability for an individual to have at least one unobserved SNP in a haplotype). Our scenarios of 5.3% and 15% missing genotypes correspond to their 15% and 39% missingness respectively.

Table 2. Performance of log-linear model under Carayol et al's (2006) Hardy-Weinberg disequilibrium scenario (Table VIII of Carayol et al 2006) using 500 replicates of 1000 triads

Pattern of missingness *	Disequilibrium coefficient	Frequency of genotypes with 0,1, and 2 copies of the risk haplotypes	Geometric mean of estimated R_1 (95% CI)	Geometric mean of estimated R_2 (95% CI)	Empirical coverage of nominal 95% confidence interval for (R1, R2)
No missing	0.1	(0.46,0.28,0.26)	1.50(1.49,1.51)	4.02(3.97,4.06)	0.952
20% missing		(0.46,0.28,0.26)	1.51(1.50,1.52)	4.03(3.98,4.08)	0.959
No missing	-0.1	(0.26,0.68,0.06)	1.50(1.450,1.51)	4.01(3.98,4.04)	0.948
20% missing		(0.26,0.68,0.06)	1.51(1.50,1.52)	4.01(3.98,4.04)	0.946

Note: Simulation conditions is Model 1 of Carayol et al. A total of 6 haplotypes comprised of 3 SNPs; risk haplotype frequency=0.4
* Our 20% missing genotypes correspond to Carayol et al's 49% missingness.

Reference:

Carayol J, Philippi A, Tores F. 2006. Estimating haplotype relative risks in complex disease from unphased SNPs data in families using a likelihood adjusted for ascertainment. Genet Epidemiol 30(8):666-76