Supplementary information:

model 1			*% of missing data*							
		expected	*0%*	*1%*	*5%*	*10%*	*20%*	*30%*	*40%*	*50%*
	number of complete families		500	486.4	434.4	374.7	271.9	189.8	125.8	77.8
without multiple imputation	*GRR2*	1	0.006	0.006	0.007	0.008	0.008	0.001	0.010	0.040
	GRR3	1.5	0.007	0.006	0.006	0.006	0.007	-0.009	-0.005	0.053
	coverage 2	0.95	0.938	0.940	0.930	0.946	0.936	0.942	0.966	0.960
	coverage 3	0.95	0.928	0.924	0.926	0.940	0.948	0.958	0.954	0.966
with multiple imputation	*GRR2*	1	0.006	0.007	0.012	0.018	0.033	0.055	0.096	0.154
	GRR3	1.5	0.007	0.008	0.013	0.015	0.022	0.027	0.036	0.045
	coverage 2	0.95	0.938	0.936	0.938	0.934	0.928	0.932	0.906	0.843
	coverage 3	0.95	0.928	0.928	0.936	0.936	0.928	0.922	0.926	0.896

model 2			*% of missing data*							
		expected	*0%*	*1%*	*5%*	*10%*	*20%*	*30%*	*40%*	*50%*
	number of complete families		500	486.2	434.3	374.6	272.1	189.5	125.2	78.4
without multiple imputation	*GRR2*	1.5	0.008	0.007	0.010	0.014	0.017	0.019	0.015	0.008
	GRR3	1.5	-0.001	-0.002	0.000	0.006	0.004	0.006	-0.006	0.000
	coverage 2	0.95	0.960	0.958	0.960	0.956	0.952	0.962	0.976	0.958
	coverage 3	0.95	0.962	0.952	0.958	0.960	0.952	0.948	0.962	0.952
with multiple imputation	*GRR2*	1.5	0.008	0.009	0.014	0.020	0.036	0.054	0.084	0.129
	GRR3	1.5	-0.001	0.001	0.006	0.012	0.017	0.023	0.021	0.018
	coverage 2	0.95	0.960	0.954	0.952	0.952	0.950	0.948	0.942	0.910
	coverage 3	0.95	0.962	0.958	0.954	0.968	0.958	0.960	0.928	0.938

model 3			*% of missing data*							
		expected	*0%*	*1%*	*5%*	*10%*	*20%*	*30%*	*40%*	*50%*
	number of complete families		500	486.7	434.3	374.6	271.6	188.6	126.4	77.6
without multiple imputation	*GRR2*	1.5	0.009	0.008	0.007	0.007	0.015	0.016	0.001	0.045
	GRR3	1.5	0.003	0.002	0.004	0.005	0.011	0.014	0.023	0.096
	coverage 2	0.95	0.922	0.924	0.912	0.938	0.940	0.942	0.932	0.964
	coverage 3	0.95	0.934	0.930	0.932	0.930	0.938	0.940	0.940	0.954
with multiple imputation	*GRR2*	1.5	0.009	0.010	0.015	0.023	0.040	0.068	0.116	0.179
	GRR3	1.5	0.003	0.005	0.009	0.015	0.022	0.032	0.049	0.058
	coverage 2	0.95	0.922	0.920	0.916	0.922	0.924	0.918	0.898	0.830
	coverage 3	0.95	0.934	0.934	0.938	0.936	0.924	0.930	0.922	0.904

model 4			% of missing data							
		expected	*0%*	*1%*	*5%*	*10%*	*20%*	*30%*	*40%*	*50%*
	number of complete families		500	486.1	433.6	375.2	272.9	189.6	125.6	78.1
without multiple imputation	*GRR2*	1.5	0.012	0.013	0.009	0.016	0.017	0.039	0.070	0.285
	GRR3	1.5	-0.002	0.000	-0.007	-0.001	0.006	0.010	0.056	0.247
	coverage 2	0.95	0.946	0.958	0.962	0.974	0.956	0.942	0.950	0.966
	coverage 3	0.95	0.958	0.956	0.956	0.956	0.960	0.970	0.968	0.952
with multiple imputation	*GRR2*	1.5	0.012	0.014	0.016	0.025	0.036	0.050	0.084	0.137
	GRR3	1.5	-0.002	0.000	0.003	0.010	0.014	0.008	0.015	0.033
	coverage 2	0.95	0.946	0.944	0.954	0.954	0.964	0.956	0.948	0.922
	coverage 3	0.95	0.958	0.954	0.964	0.960	0.970	0.958	0.948	0.934

Supplementary table 1: Bias and coverage of genetic parameter estimates from one-locus simulation study as a function of the percentage of missing data. The average number of informative families for the analysis without MI is noted on the third row.

	haplotype frequencies				number of families
	11	**12**	**21**	**22**	
population1	0.5	0.1	0.25	0.15	100
population 2	0.2	0.3	0.3	0.2	400

Supplementary table 2: Haplotype frequencies for simulations under population stratification

Supplementary Figure 1: LD pattern between the five loci considered in the simulations. Gradation of greys represents the level of the r^2 and numbers are the D' values. This LD pattern is similar to the one observed in the CTLA4 gene in a sample of 450 multiple sclerosis trios [23].

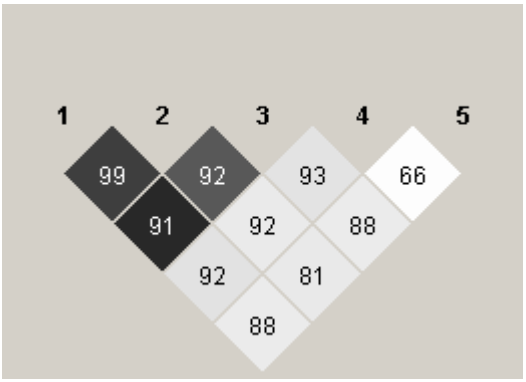
Supplementary Figure 2: Correlation plot between the logarithm of the p-value obtained on the complete data file and data with 5, 30 and 50% of missing data when using MI approach on the recessive model 1 (a, b and c respectively) and the dominant model 2 (d, e and f) with genotype relative risks 1.5. For a better understanding of these graphs, the linear equation y=x has been plotted.

Supplementary Figure 3: Correlation plot between the genotype relative risk of the homozygous 2/2 obtained on the complete data file and data with 5, 30 and 50% of missing data when using MI approach on the recessive model 1 (a, b and c respectively) and the dominant model 2 (d, e and f) with genotype relative risk 1.5. For a better understanding of these graph, the linear equation y=x has been plotted.

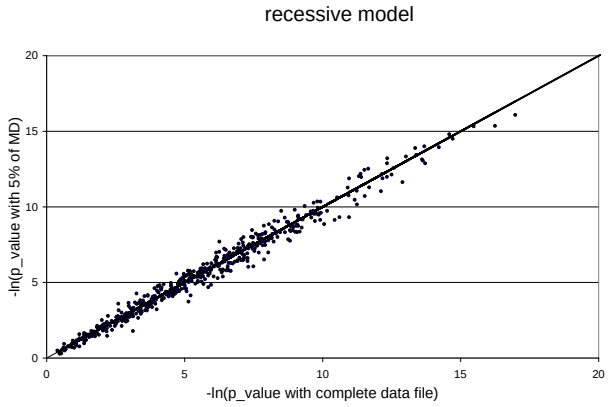
Supplementary Figure 4: mean difference between the true p-value from complete datasets and p-values obtained by TRANSMIT or conditional logistic regression when using or not MI approach from datasets with different levels of missing data.

Supplementary Figure 5: Type one error rate obtained at α=0.05 on data simulated under population stratification as a function of the percentage of missing data when using TRANSMIT program or conditional logistic regression with or without MI.

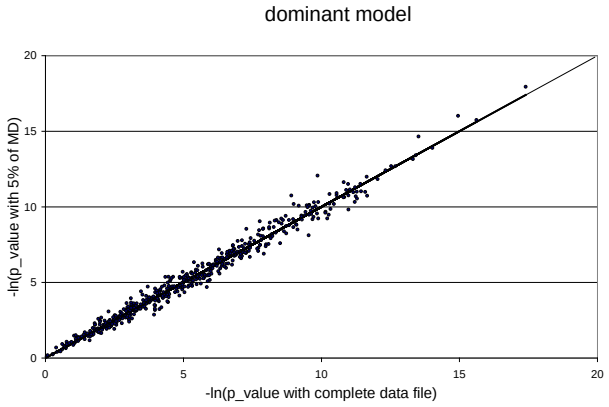
Supplementary information, figure1



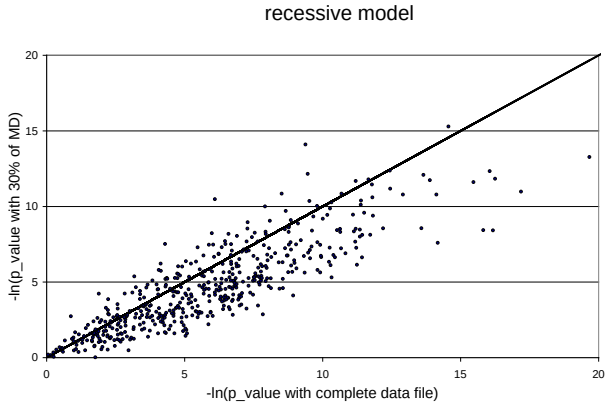
a.



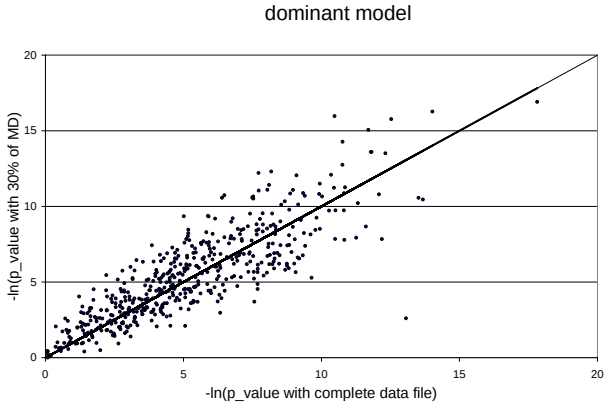
d.



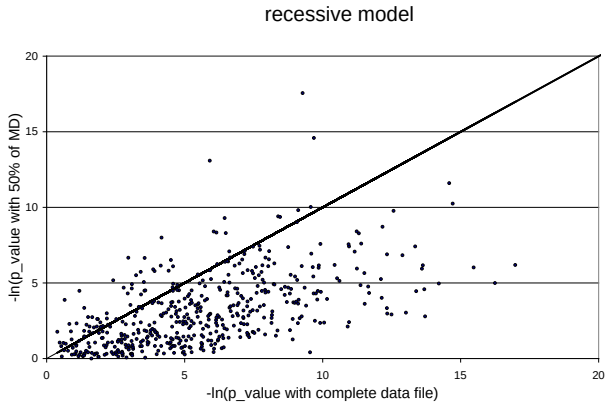
b.



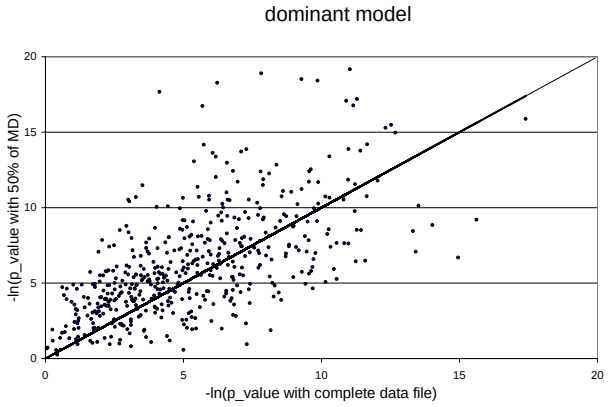
e.



c.



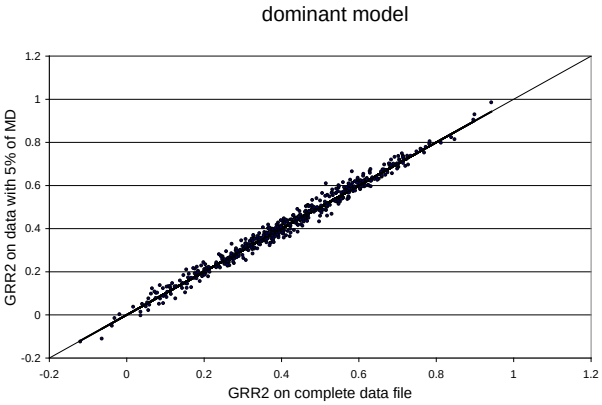
f.



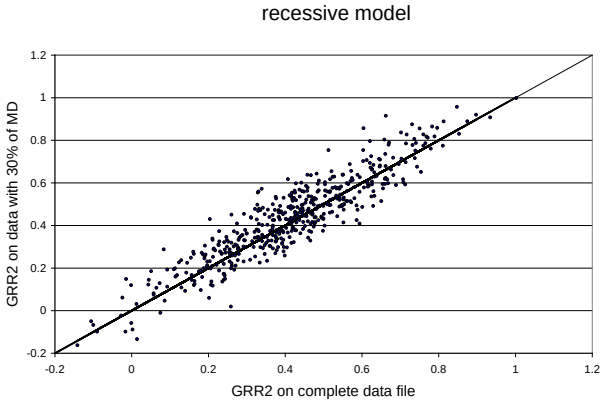
Supplementary information, figure 2

a

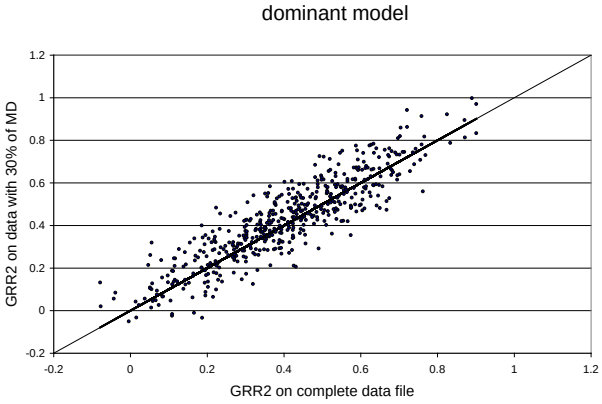
d



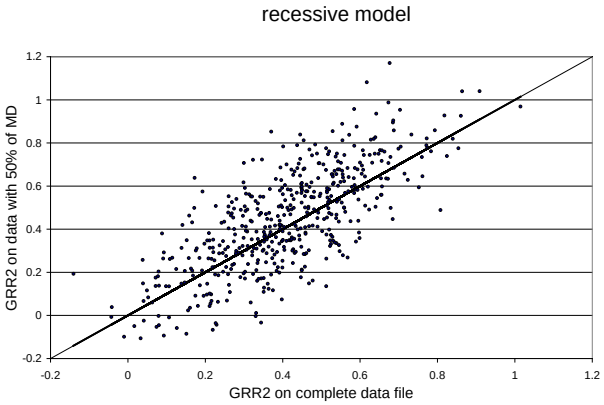
b



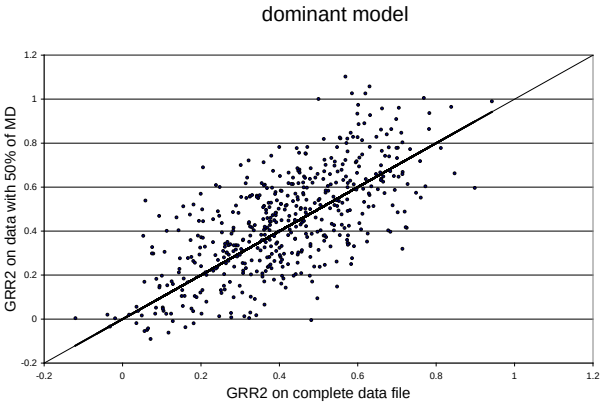
e



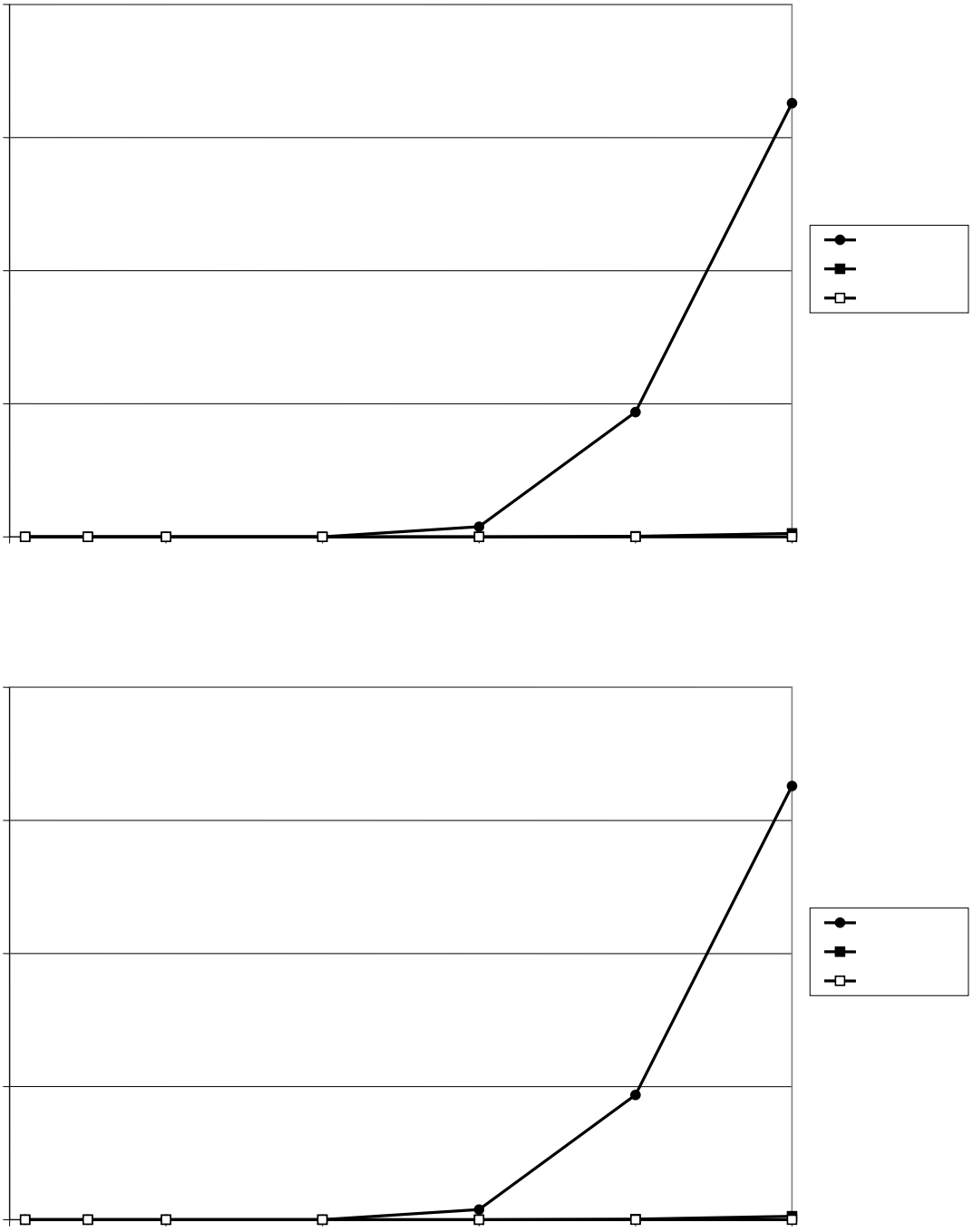
c



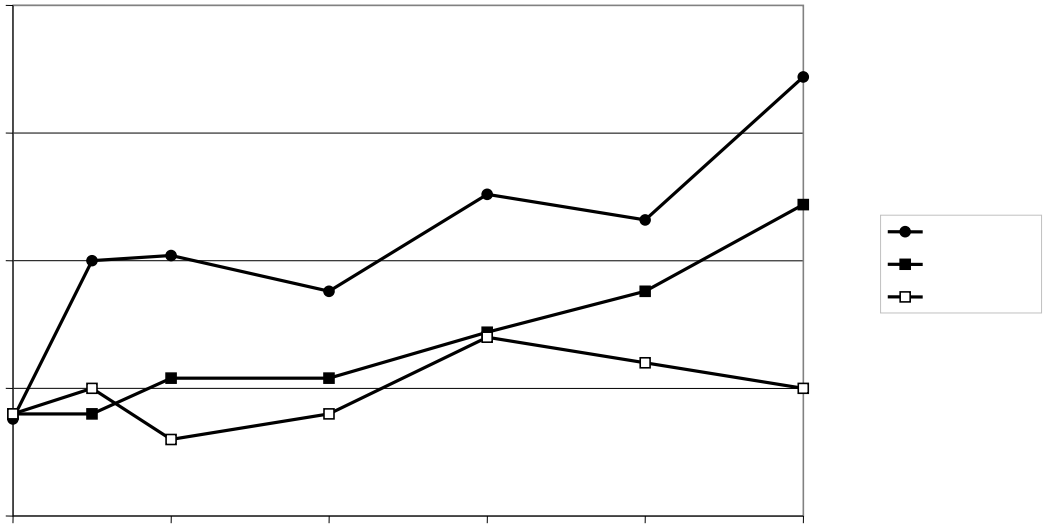
f



Supplementary information, figure 3



Supplementary information, figure 4



Supplementary information, figure 5