

## 1. Supplementary file for simulations studying the effect of sample size on the methods

This supplementary file contains more detailed results related to the simulation study presented in subsection 5.2. The goal of the simulation study in subsection 5.2 was to understand the effect of sample size on the three approaches considered in this paper based on QRR, Lasso and rank-ordering (RO) of univariate p-values. Details of the simulation data generation were presented in the main text (subsection 5.2) as well as in Appendix C.

There were two overarching simulations designs: one with 1000 SNPs and a second design based on 10,000 SNPs. In both cases, between-SNP correlations were incorporated in blocks of 10 SNPs, with two correlation ranges (0.05, 0.15) and (0.40, 0.60) yielding two separate simulation scenarios. For the design based on 1000 SNPs, the sample sizes considered were 300, 650, 1000, 2000 and 3000. The only sample size considered for the design with 10,000 SNPs, was ( $N =$ ) 3000. Thus, based on the ratio of number of SNPs to sample size, the design with 10,000 SNPs and 3000 samples can be considered similar to the design with 1000 SNPs and 300 samples. The number of true underlying shared SNPs incorporated in all scenarios with 1000 SNPs was ( $M_c =$ ) 100, and in scenarios with 10,000 SNPs was ( $M_c =$ ) 500.

In order to save space in the main text, only summarized results pooled across all simulations at a  $\lambda$ -threshold yielding approximately 100 shared SNPs, were presented in subsection 5.2 (see Table 1). However, originally (non-summarized) results were available for  $\lambda$ -thresholds corresponding to the whole range of shared subset sizes. The purpose of this supplementary file is to present these more detailed results, using plots. Note also that in the main text, we presented only results that corresponded to the design with 1000 SNPs. The plots presented in this file also include those corresponding to the design with 10,000 SNPs. Table 1 from subsection 5.2 is repeated at the end of this file, but now including results from design with 10,000 SNPs also.

Although this file contains more detailed results, the conclusions are essentially the same as that presented in subsection 5.2: the performance of all the methods improved with increase in sample size, not very surprisingly. Consistent with results in the subsection 5.1, the performance of all methods were dramatically larger when there was more correlation present among the SNPs. Although all these methods show dramatic increase in performance with higher SNP-correlations, the one with the biggest improvement was QRR. It is interesting to note that the performance of the methods with a sample size of 300 with higher between-SNP correlations was nearly the same as with a sample size of 3000 but with low correlations.

More interesting to note is that in the low correlation scenario with  $N \leq M$  (i.e.  $N = 300, 650$  and  $1000$ ) the rank-ordering (RO) method outperformed the other two methods and Lasso had a slight edge over QRR. In the low correlation cases with  $N > M$  (i.e.  $N = 2000$  and  $3000$ ), QRR is better than Lasso but still underperforming compared to RO approach. In the high correlation scenarios,  $RO > QRR > \text{Lasso}$  performance-wise for  $N = 300$  only, and  $QRR > RO > \text{Lasso}$  for all other  $N$  in Table 1 (i.e.  $N \geq 650$ ).

## 2. Plots from simulations for studying the effect of sample size

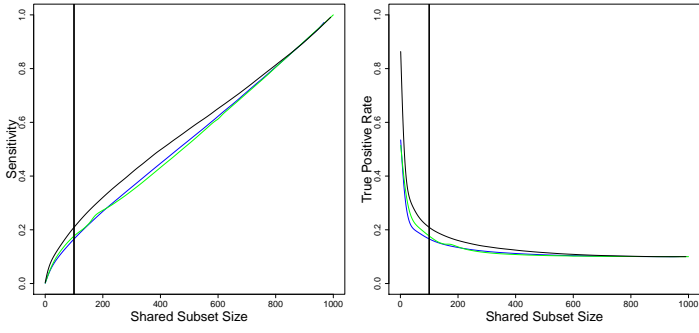
The results are plotted in figures 1 to 4. Figures 1 and 2 plot the results for between-SNP correlation range (0.05-0.15), and figures 3 and 4 plot the results for the correlation range (0.40-0.60). Figures 1 and 3 are the main figures. Figures 2 and 4 essentially repeat some of the information presented in figures 1 and 3, respectively, but now narrowing the focus to a smaller range of shared-subset-size on the  $x$ -axis, and plotting the pooled results with 95% confidence intervals, for bins of shared subsets. Thus the results in Table 1, essentially corresponds to one bin in figures 2 and 4.

All figures contain six subplots labelled (a) to (f). Within each subplot, there are two panels, a left panel and a right panel. The left panel plots sensitivity against shared subset size detected at a given  $\lambda$ -threshold, while as the right panel plots true positive rate against the shared subset size. Within each panel the blue curve corresponds to QRR, green corresponds to Lasso and black corresponds to RO. A vertical line is plotted at the  $x$ -axis value corresponding to a shared subset size of 100.

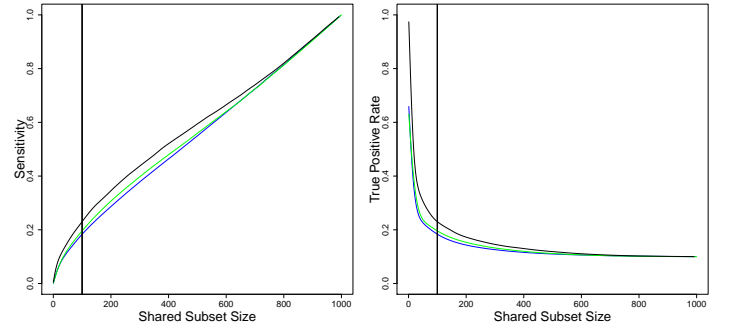
The first five subplots ((a)-(e)) in each figure correspond to the simulation design with 1000 SNPs (i.e.  $M$ ) and 100 underlying true shared SNPs (i.e.  $M_c$ ). The last subplot (f) in each figure plots the results for the simulation design with 10,000 SNPs and 500 underlying true shared SNPs. Subplots (a) to (e) correspond to sample sizes of 300, 650, 1000, 2000 and 3000, respectively. As mentioned previously, only one sample size ( $N =$ ) 3000 was considered for the design with 10,000 SNPs. The results are based on 1000 iterations in the design scenarios with 1000 SNPs, and 100 iterations for the design with 10,000 SNPs. The curves plotted in all figures were obtained from the loess scatter-smooth plotter applied across the curves from all iterations.

Since the  $\lambda$ -grid that we chose was fine enough and large enough, figures 1 and 3 plot the results for the entire range of possible shared subset sizes that can be selected - that is, 0-1000 for subplots (a) to (e) and 0-10000 for the subplot (f). In figures 2 and 4, we narrow the focus to shared subset size range of 0-300 because in practice selecting a shared subset of size larger than 300 is often not feasible.

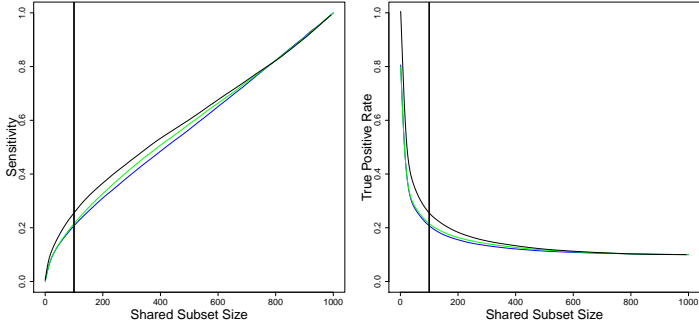
For the convenience of the reader, all the plots in figures 1 to 4, but with enlarged versions, are also presented in a separate supplementary file labelled ‘Suppl-file-sample-size-sim-B.pdf’.



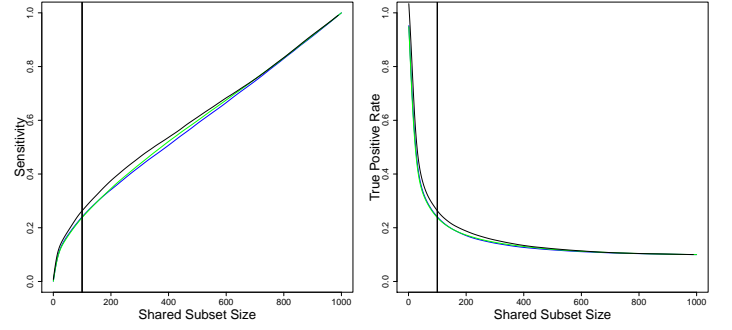
(a)  $M = 1000, M_c = 100, N = 300$



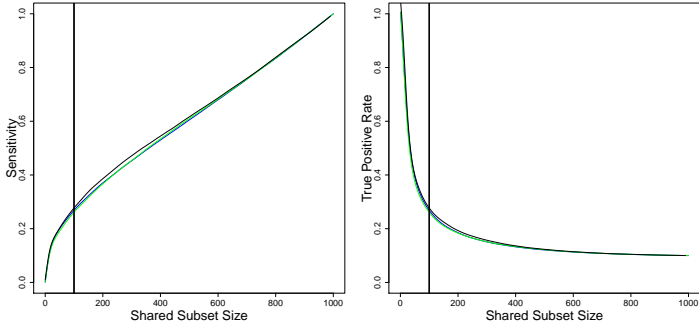
(b)  $M = 1000, M_c = 100, N = 650$



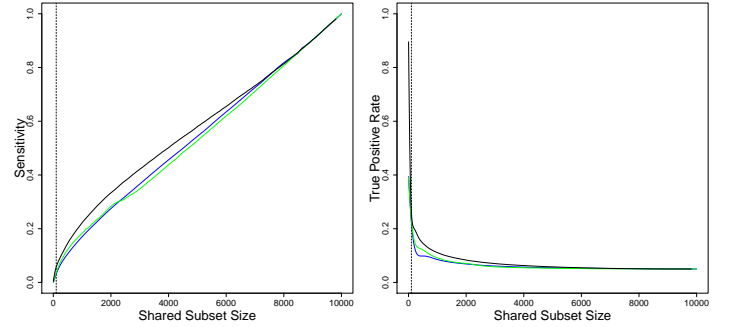
(c)  $M = 1000, M_c = 100, N = 1000$



(d)  $M = 1000, M_c = 100, N = 2000$

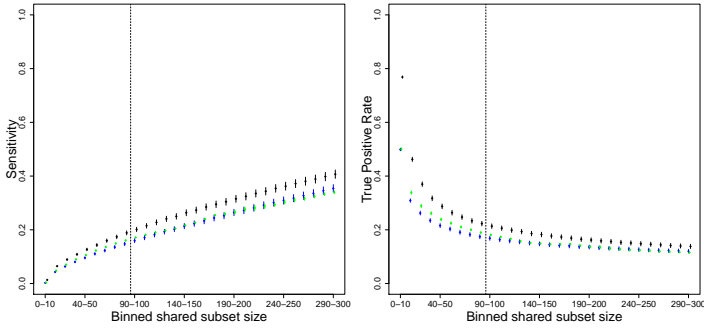


(e)  $M = 1000, M_c = 100, N = 3000$

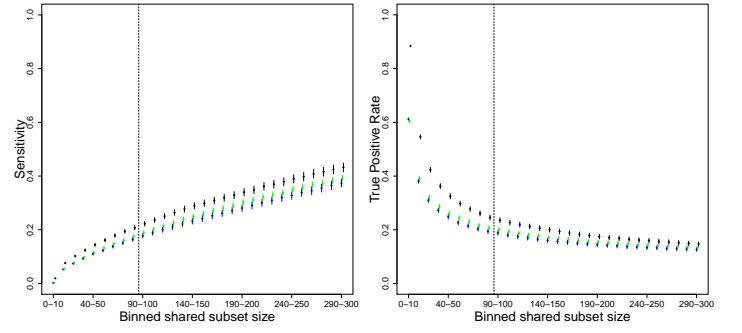


(f)  $M = 10000, M_c = 500, N = 3000$

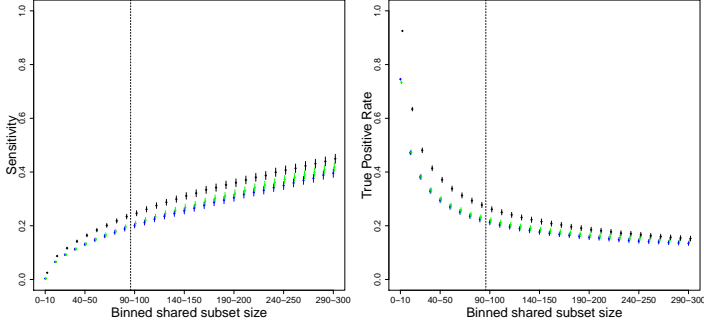
Figure 1: Effect of sample size on performance of the methods. QRR: blue, Lasso: green, RO: black. Between-SNP correlation range: 0.05-0.15.  $x$ -axis in each panel plots the entire range of possible shared subsets that may be selected. See pages 1-2 for more description.



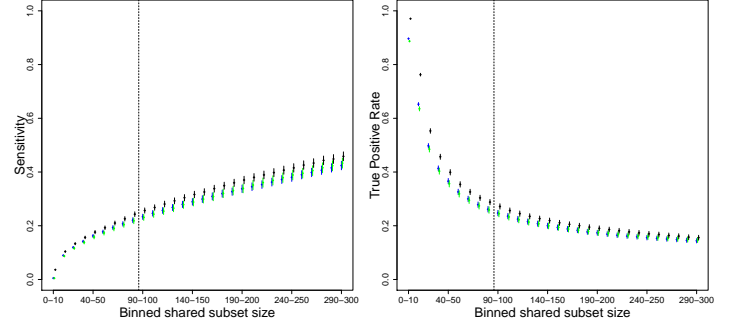
(a)  $M = 1000, M_c = 100, N = 300$



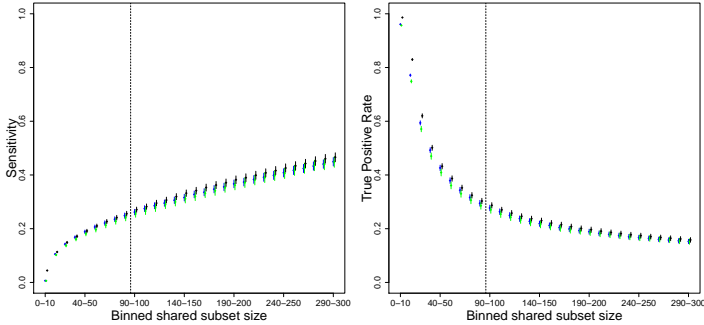
(b)  $M = 1000, M_c = 100, N = 650$



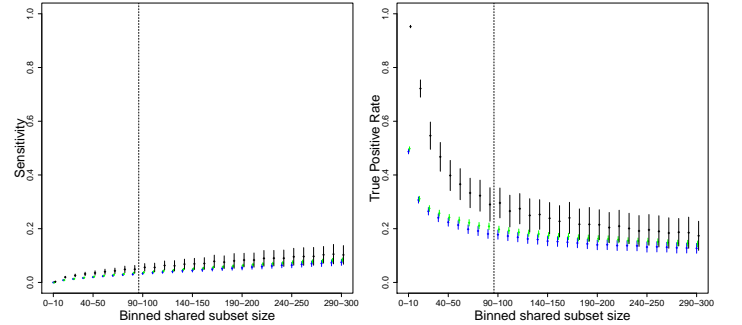
(c)  $M = 1000, M_c = 100, N = 1000$



(d)  $M = 1000, M_c = 100, N = 2000$

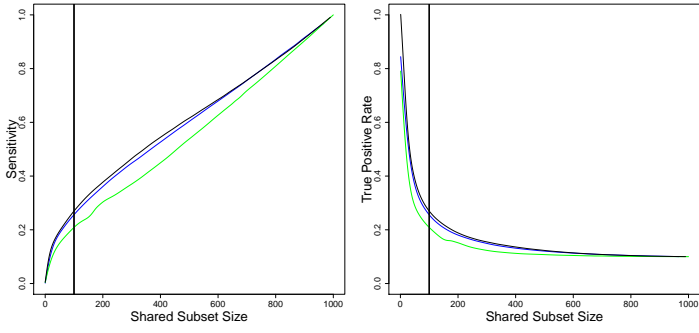


(e)  $M = 1000, M_c = 100, N = 3000$

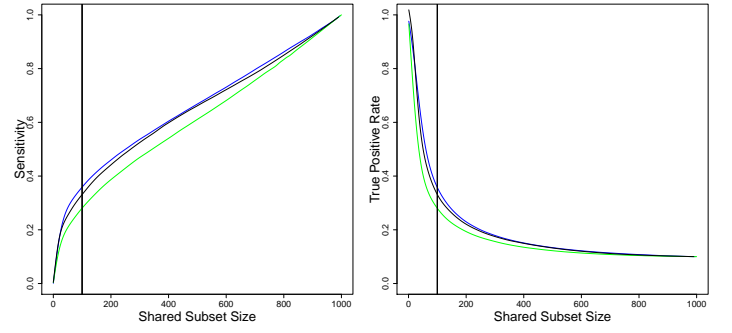


(f)  $M = 10000, M_c = 500, N = 3000$

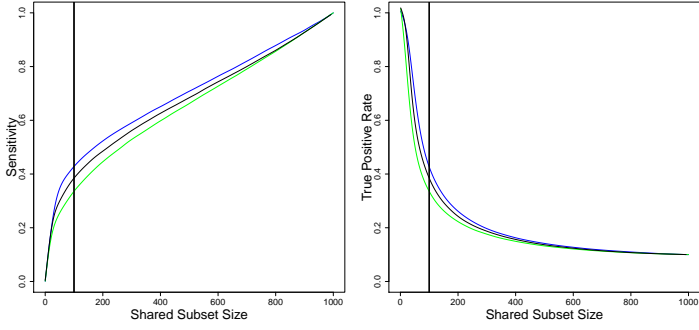
Figure 2: Effect of sample size on performance of the methods. QRR: blue, Lasso: green, RO: black. Between-SNP correlation range: 0.05-0.15.  $x$ -axis in each panel corresponds to bins of size 10 for shared subsets. Thus this figure presents the same information as in Figure 1, but pooled within each bin. The range of shared subset sizes on the  $x$ -axis is limited to 0-100. The vertical bars correspond to 95% confidence intervals. In figure (f) the vertical bars are longer because only 100 iterations were used. See pages 1-2 for more description.



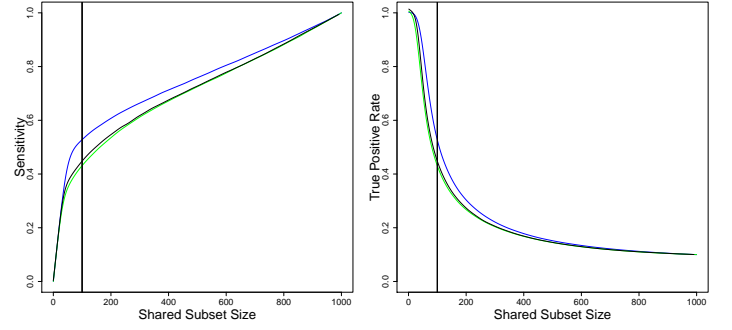
(a)  $M = 1000, M_c = 100, N = 300$



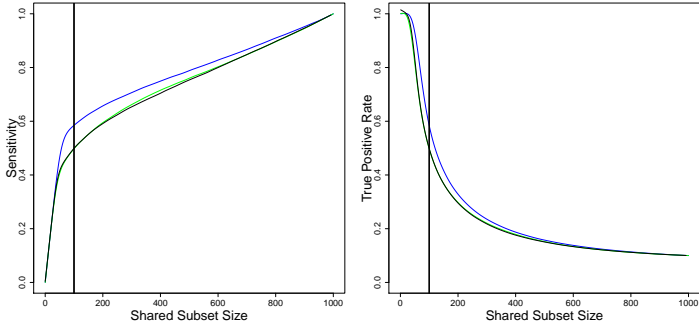
(b)  $M = 1000, M_c = 100, N = 650$



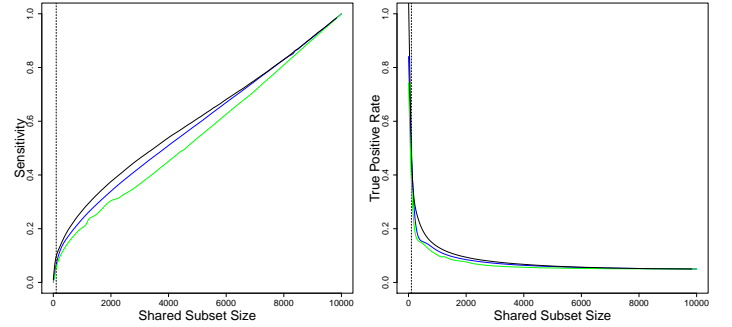
(c)  $M = 1000, M_c = 100, N = 1000$



(d)  $M = 1000, M_c = 100, N = 2000$

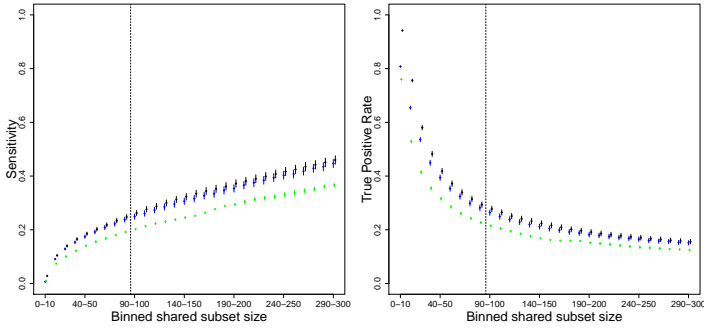


(e)  $M = 1000, M_c = 100, N = 3000$

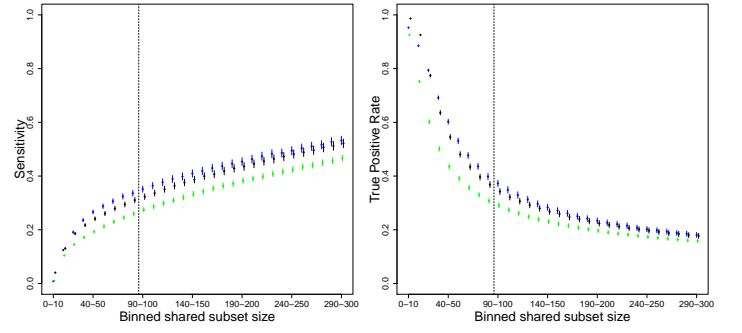


(f)  $M = 10000, M_c = 500, N = 3000$

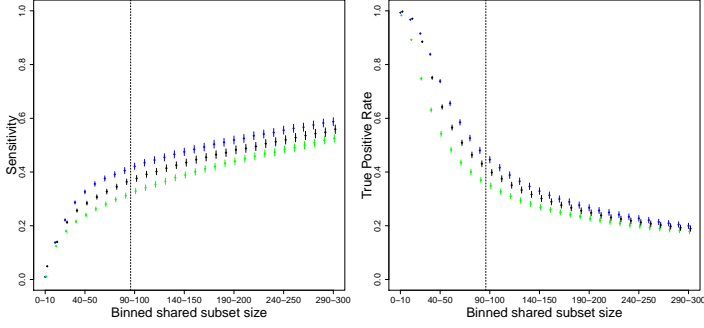
Figure 3: Effect of sample size on performance of the methods. QRR: blue, Lasso: green, RO: black. Between-SNP correlation range: 0.40-0.60.  $x$ -axis in each panel plots the entire range of possible shared subsets that may be selected. See pages 1-2 for more description.



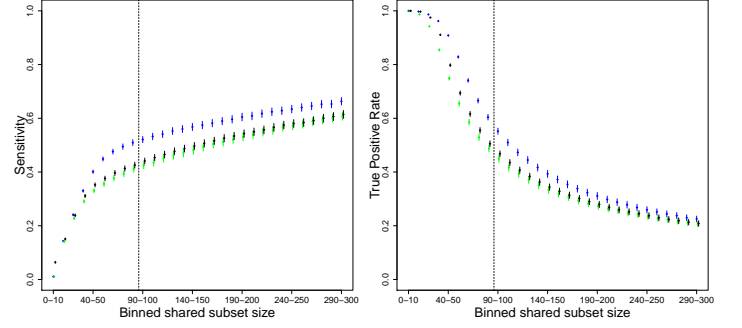
(a)  $M = 1000, M_c = 100, N = 300$



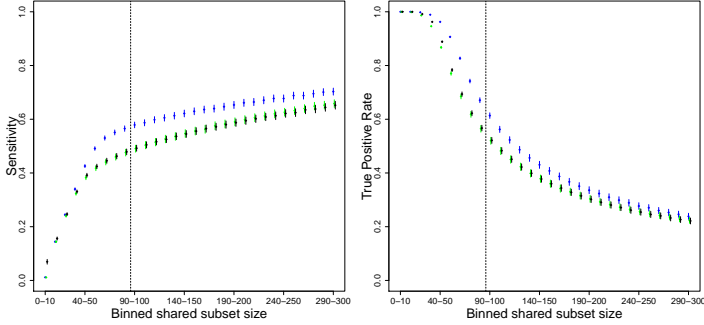
(b)  $M = 1000, M_c = 100, N = 650$



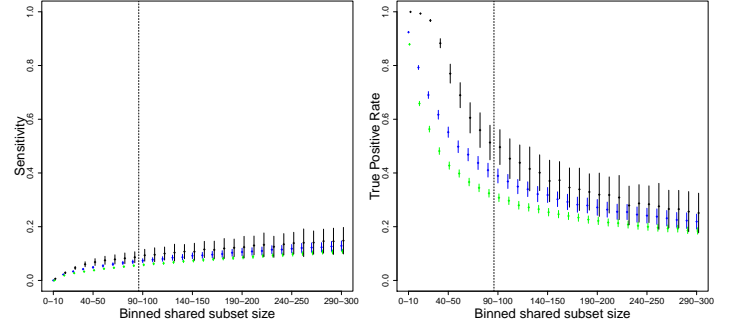
(c)  $M = 1000, M_c = 100, N = 1000$



(d)  $M = 1000, M_c = 100, N = 2000$



(e)  $M = 1000, M_c = 100, N = 3000$



(f)  $M = 10000, M_c = 500, N = 3000$

Figure 4: Effect of sample size on performance of the methods. QRR: blue, Lasso: green, RO: black. Between-SNP correlation range: 0.40-0.60.  $x$ -axis in each panel corresponds to bins of size 10 for shared subsets. Thus this figure presents the same information as in Figure 1, but pooled within each bin. The range of shared subset sizes on the  $x$ -axis is limited to 0-100. The vertical bars correspond to 95% confidence intervals. In figure (f) the vertical bars are longer because only 100 iterations were used. See pages 1-2 for more description.

### 3. Table 1, repeated

Table 1 in subsection 5.2 included only results corresponding to the simulation design with 1000 SNPs. We present the Table 1 again here, but now including results from design with 10,000 SNPs also.

It is illuminating to compare the results from the design with 10,000 SNPs and 3000 sample size, to that with 1000 SNPs and 300 sample size. The sensitivity values in Table 1 below corresponding to 10,000 SNPs is quite low, but that is to be expected since it is obvious that a shared subset of size 100 is too small in this case. A more meaningful comparison can be based on TPR values.

The TPR values are higher in the 10,000 SNPs design case perhaps because the sample size is also higher. The increase in TPR values is only nominal for QRR and Lasso when the between-SNPs correlation range is (0.05-0.15), but is more substantial for the RO method. This result makes sense since RO is a univariate approach and hence the increase in sample size helps irrespective of the fact that the number of SNPs also increased. When the correlation range is (0.40-0.60), all three methods show substantial improvement in TPR in the 10,000 SNPs case compared to 1000 SNPs case. However even in this case, RO outperforms the other two, again perhaps because the larger sample size helps the univariate method a lot more than the multivariate methods. Among the multivariate approaches, QRR's performance is better than of Lasso when the correlation range is (0.40-0.60) with 10,000 SNPs.

Sample Size						
$M$	$M_c$	$\rho$ , range	$N$	Method	Sensitivity*	True Pos. Rate*
1000	100	0.05-0.15	300	QRR	0.171 (0.163, 0.179)	0.163 (0.156, 0.171)
				Lasso	0.181 (0.178, 0.184)	0.173 (0.171, 0.176)
				Rank Ord.	0.215 (0.206, 0.224)	0.206 (0.198, 0.214)
			650	QRR	0.189 (0.180, 0.197)	0.180 (0.172, 0.189)
				Lasso	0.202 (0.194, 0.210)	0.193 (0.186, 0.201)
				Rank Ord.	0.237 (0.227, 0.246)	0.227 (0.218, 0.236)
			1000	QRR	0.212 (0.202, 0.221)	0.203 (0.194, 0.212)
				Lasso	0.221 (0.211, 0.231)	0.212 (0.202, 0.221)
				Rank Ord.	0.261 (0.251, 0.271)	0.251 (0.240, 0.260)
			2000	QRR	0.233 (0.223, 0.243)	0.247 (0.237, 0.257)
				Lasso	0.231 (0.220, 0.240)	0.245 (0.233, 0.256)
				Rank Ord.	0.257 (0.247, 0.266)	0.272 (0.262, 0.282)
			3000	QRR	0.262 (0.252, 0.272)	0.277 (0.267, 0.288)
				Lasso	0.255 (0.243, 0.267)	0.270 (0.258, 0.283)
				Rank Ord.	0.271 (0.261, 0.280)	0.287 (0.277, 0.297)
1000	100	0.40-0.60	300	QRR	0.261 (0.250, 0.271)	0.250 (0.240, 0.260)
				Lasso	0.214 (0.210, 0.217)	0.205 (0.201, 0.208)
				Rank Ord.	0.277 (0.266, 0.287)	0.265 (0.255, 0.275)
			650	QRR	0.364 (0.352, 0.376)	0.349 (0.337, 0.361)
				Lasso	0.286 (0.277, 0.295)	0.274 (0.266, 0.282)
				Rank Ord.	0.337 (0.326, 0.348)	0.322 (0.311, 0.333)
			1000	QRR	0.435 (0.422, 0.447)	0.416(0.404, 0.429)
				Lasso	0.342 (0.331, 0.352)	0.327 (0.316, 0.337)
				Rank Ord.	0.391 (0.380, 0.403)	0.375 (0.363, 0.386)
			2000	QRR	0.532 (0.520, 0.545)	0.510 (0.498, 0.522)
				Lasso	0.435 (0.422, 0.448)	0.417 (0.404, 0.429)
				Rank Ord.	0.454 (0.442, 0.465)	0.435 (0.423, 0.447)
			3000	QRR	0.579 (0.568, 0.589)	0.614 (0.603, 0.624)
				Lasso	0.490 (0.478, 0.503)	0.520 (0.507, 0.533)
				Rank Ord.	0.492 (0.481, 0.503)	0.521 (0.510, 0.532)
10,000	500	0.05-0.15	3000	QRR	0.036 (0.032, 0.040)	0.173 (0.156, 0.190)
				Lasso	0.040 (0.037, 0.043)	0.193 (0.181, 0.205)
				Rank Ord.	0.056 (0.040, 0.072)	0.266 (0.207, 0.325)
		0.40-0.60		QRR	0.077 (0.069, 0.085)	0.368 (0.342, 0.395)
				Lasso	0.062 (0.058, 0.066)	0.297 (0.282, 0.311)
				Rank Ord.	0.096 (0.070, 0.121)	0.453 (0.379, 0.527)
*Means and 95% CIs (in parenthesis) for shared subset sizes between 90 and 100, pooled across all iterations						