Supplementary material for "Sequence kernel association analysis of rare variant set based on the marginal regression model for binary traits"

Baolin Wu[1], James S. Pankow[2], Weihua Guan[1]

[1]Division of Biostatistics, [2]Division of Epidemiology and Community Health

School of Public Health, University of Minnesota

1 Simulation studies

We conducted extensive simulation studies to evaluate the performance of the proposed and existing methods. Following Lee *et al.* (2012), we generated 10,000 European-like haplotypes of length 1000 kb under a calibrated coalescent model (Schaffner *et al.*, 2005). We randomly pair the haplotypes to simulate a total population of 10^6 individuals. We randomly select a gene region of length 10 kb and study those rare variants with MAF≤ 0.01. We consider two covariates $Z = (Z_1, Z_2)'$: $Z_1 \in \{0, 1\}$ follows Bernoulli(0.5), and $Z_2 \sim N(0, 1)$. We model the logit disease risk as expit$(\beta_0 + Z'\beta_Z + \sum_{j=1}^{m} \beta_j G_j)$. We set $\beta_0 = -3.4, \beta_Z = (0.5, 0.5)'$ (corresponding to 5% population disease rate). We randomly select n_e cases and n_c controls from the simulated population of 10^6 samples. We compared five rare variant set analysis methods: SKAT, SKAT-O, SKAT$_L$, SKAT-O$_L$ and burden test. In the burden and SKAT tests, we assign weight Beta$(p_j; a_0, b_0)$ to the jth variant G_j. And for the proposed method we assign weight Beta$(p_j; a_1, b_1)$. Here p_j is the MAF of G_j. For a given variant, the likelihood ratio test statistic is inherently standardized and roughly corresponds to the standardized score statistics, which

Table 1: Type I error of rare variant set analysis: $n_e = 2500$ cases and $n_c = 2500$ controls. The Type I errors have been divided by their nominal levels

(a_1, b_1)	(1,25)			(1.5,25.5)			(2,26)		
α	10^{-4}	10^{-3}	10^{-2}	10^{-4}	10^{-3}	10^{-2}	10^{-4}	10^{-3}	10^{-2}
SKAT_L	0.69	0.74	0.90	1.07	1.05	1.04	1.06	1.02	1.03
SKAT-O_L	0.91	0.92	0.99	0.96	1.00	1.05	1.10	1.12	1.07

Table 2: Type I error of rare variant set analysis: $n_e = 1700$ cases and $n_c = 3300$ controls. The Type I errors have been divided by their nominal levels

(a_1, b_1)	(1,25)			(1.5,25.5)			(2,26)		
α	10^{-4}	10^{-3}	10^{-2}	10^{-4}	10^{-3}	10^{-3}	10^{-4}	10^{-3}	10^{-2}
SKAT_L	0.75	0.83	0.95	1.08	1.07	1.05	0.93	1.06	1.05
SKAT-O_L	0.93	0.95	1.03	0.92	0.98	1.02	1.12	1.12	1.09

is the score statistics used in SKAT scaled by its standard error, which is roughly proportional to $\sqrt{p_j(1-p_j)}$. Therefore for the proposed method, we set $a_1 = a_0 + 0.5$ and $b_1 = b_0 + 0.5$. We investigated three sets of weights for (a_0, b_0): (0.5,24.5), (1,25), and (1.5,25.5); and two case-control ratios for (n_e, n_c): (2500,2500) and (1700,3300).

We use 10^7 experiments to evaluate the type I error at the nominal significance level $\alpha = 10^{-4}, 10^{-3}$, and 10^{-2} by setting all $\beta_j = 0$. The results are summarized in Table 1 and 2 for the two case-control ratios. We can see that the the proposed methods appropriately controlled the Type I errors. The Type I errors are generally protected under different weights, which is consistent with the observations of Wu *et al.* (2011).

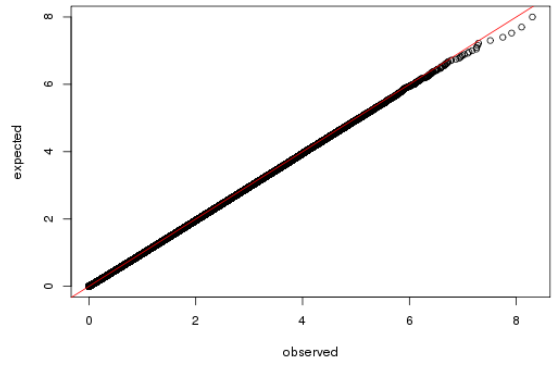
We also conducted 10^8 simulations to verify that the proposed methods have well-calibrated p-values at very stringent genome-wide significance level. Table 3 shows that the proposed methods appropriately control type I errors at small α levels. Figure 1 shows the corresponding QQ plots for the proposed SKAT_L and SKAT-O_L with weights $(a_1 = 1.5, b_1 = 25.5)$ based on 10^8 simulations.

We use 10^4 experiments to evaluate the power under various combinations of β_j at $\alpha = 10^{-6}, 10^{-5}, 10^{-4}$, and 10^{-3}. The rare variant effects β_j are set as follows. Each time we randomly select θ proportion of rare variants and set their $|\beta_j| = d \log_{10}(p_j)$. The other null rare variants have zero coefficients. We have assumed that rarer variants have larger effect sizes. We conducted

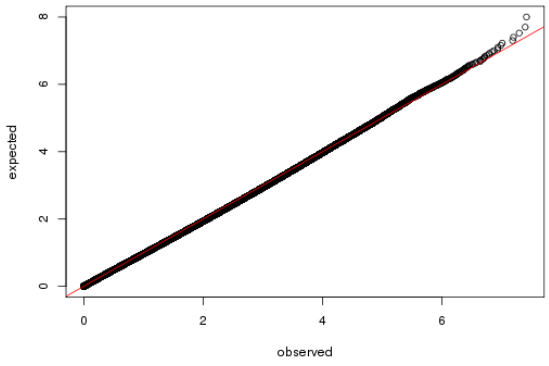
Table 3: Type I error estimates of proposed methods with weight $(a_1 = 1.5, b_1 = 25.5)$ at stringent α level based on 10^8 simulations: Type I error has been scaled by the nominal significance level α.

(n_e, n_c)	(2500,2500)		(1700,3300)	
α	10^{-5}	10^{-6}	10^{-5}	10^{-6}
SKAT_L	1.08	1.06	1.05	1.09
SKAT-O_L	0.93	0.89	0.93	0.78

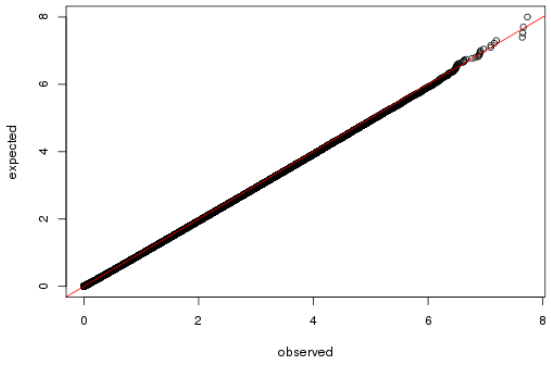
Figure 1: QQ plots of 10^8 p-values: plotted are the negative log (base 10) transformed p-values vs their expected values.



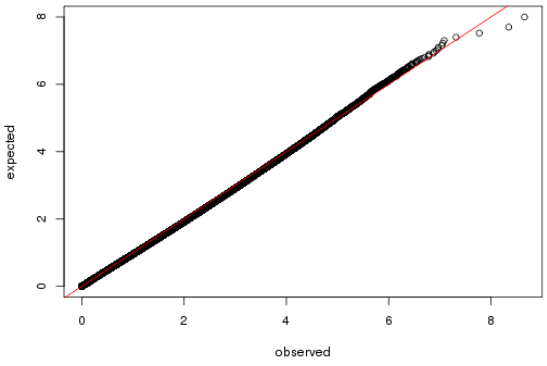
(a) SKAT_L p-values: $n_e = n_c = 2500$



(b) SKAT-O_L p-values: $n_e = n_c = 2500$



(c) SKAT_L p-values: $n_e = 1700, n_c = 3300$



(d) SKAT-O_L p-values: $n_e = 1700, n_c = 3300$

simulations for (1) $\theta = 0.05, d = -0.6$, (2) $\theta = 0.1, d = -0.5$, (3) $\theta = 0.2, d = -0.4$, (4) $\theta = 0.5, d = -0.25$. They correspond to odds ratio of 3.32, 2.72, 2.23 and 1.65 for MAF=0.01 respectively. We consider two scenarios for the direction of causal variant effects. First, we assume a mix of equal proportions of protective and deleterious variants, which will in general favor the kernel association test. Second, we assume a mix of unequal proportions of protective and deleterious variants. Specially we randomly set signs of β_j as negative or positive with probability 0.9 and 0.1 respectively.

Tables 4 through 9 summarized the power for $n_e = n_c = 2500$. And Tables 10 through 15 summarized the power for $n_e = 1700, n_c = 3300$. Overall we can see that the proposed SKAT$_L$ and SKAT-O$_L$ have improved performance compared to the SKAT based approaches. When a large proportion of the variants are causal ($\theta = 0.5$) with a mix of protective and deleterious variants, the SKAT-O$_L$ can adapt to the data and has the overall best performance.

Table 4: Power comparison of rare variant set analysis: $n_e = n_c = 2500$, $a_0 = 1, b_0 = 25$, equal proportions of protective and deleterious variants. The highest powered tests in each row are bold-faced.

$\theta = 0.05, d = -0.6$					
α	SKAT	SKAT-O	SKAT$_L$	SKAT-O$_L$	Burden
10^{-6}	0.165	0.137	**0.2**	0.166	0.003
10^{-5}	0.228	0.2	**0.263**	0.234	0.007
10^{-4}	0.308	0.281	**0.352**	0.319	0.018
10^{-3}	0.429	0.399	**0.474**	0.44	0.045
$\theta = 0.1, d = -0.5$					
α	SKAT	SKAT-O	SKAT$_L$	SKAT-O$_L$	Burden
10^{-6}	0.247	0.204	**0.294**	0.244	0.008
10^{-5}	0.347	0.303	**0.391**	0.351	0.016
10^{-4}	0.461	0.426	**0.505**	0.469	0.035
10^{-3}	0.603	0.566	**0.645**	0.609	0.075
$\theta = 0.2, d = -0.4$					
α	SKAT	SKAT-O	SKAT$_L$	SKAT-O$_L$	Burden
10^{-6}	0.348	0.295	**0.396**	0.338	0.016
10^{-5}	0.474	0.424	**0.519**	0.47	0.03
10^{-4}	0.612	0.57	**0.651**	0.613	0.063
10^{-3}	0.758	0.728	**0.788**	0.761	0.117
$\theta = 0.5, d = -0.25$					
α	SKAT	SKAT-O	SKAT$_L$	SKAT-O$_L$	Burden
10^{-6}	0.296	0.241	**0.338**	0.278	0.014
10^{-5}	0.431	0.38	**0.472**	0.42	0.028
10^{-4}	0.587	0.543	**0.627**	0.583	0.057
10^{-3}	0.755	0.723	**0.784**	0.754	0.112

Table 5: Power comparison of rare variant set analysis: $n_e = n_c = 2500$, $a_0 = 0.5, b_0 = 24.5$, equal proportions of protective and deleterious variants. The highest powered tests in each row are bold-faced.

$\theta = 0.05, d = -0.6$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.045	0.025	**0.086**	0.05	0
10^{-5}	0.095	0.058	**0.149**	0.102	0
10^{-4}	0.182	0.134	**0.251**	0.202	0.003
10^{-3}	0.326	0.275	**0.403**	0.348	0.013
$\theta = 0.1, d = -0.5$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.123	0.078	**0.186**	0.123	0.001
10^{-5}	0.215	0.151	**0.293**	0.217	0.002
10^{-4}	0.356	0.291	**0.437**	0.366	0.008
10^{-3}	0.548	0.485	**0.62**	0.559	0.027
$\theta = 0.2, d = -0.4$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.253	0.176	**0.33**	0.239	0.003
10^{-5}	0.392	0.301	**0.469**	0.38	0.007
10^{-4}	0.56	0.484	**0.63**	0.555	0.017
10^{-3}	0.746	0.688	**0.795**	0.746	0.05
$\theta = 0.5, d = -0.25$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.205	0.141	**0.264**	0.185	0.002
10^{-5}	0.332	0.25	**0.4**	0.311	0.007
10^{-4}	0.515	0.437	**0.576**	0.501	0.02
10^{-3}	0.721	0.661	**0.764**	0.717	0.052

Table 6: Power comparison of rare variant set analysis: $n_e = n_c = 2500$, $a_0 = 1.5, b_0 = 25.5$, equal proportions of protective and deleterious variants. The highest powered tests in each row are bold-faced.

$\theta = 0.05, d = -0.6$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.113	0.104	**0.13**	0.12	0.012
10^{-5}	0.153	0.139	**0.17**	0.157	0.022
10^{-4}	0.207	0.194	**0.224**	0.212	0.042
10^{-3}	0.281	0.267	**0.302**	0.289	0.076
$\theta = 0.1, d = -0.5$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.151	0.138	**0.171**	0.158	0.02
10^{-5}	0.213	0.198	**0.237**	0.221	0.033
10^{-4}	0.296	0.282	**0.319**	0.304	0.06
10^{-3}	0.405	0.392	**0.431**	0.414	0.109
$\theta = 0.2, d = -0.4$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.188	0.178	**0.211**	0.2	0.033
10^{-5}	0.271	0.256	**0.298**	0.284	0.055
10^{-4}	0.386	0.372	**0.413**	0.397	0.088
10^{-3}	0.534	0.515	**0.558**	0.54	0.147
$\theta = 0.5, d = -0.25$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.154	0.143	**0.174**	0.163	0.025
10^{-5}	0.236	0.219	**0.254**	0.24	0.048
10^{-4}	0.353	0.338	**0.374**	0.36	0.082
10^{-3}	0.523	0.504	**0.545**	0.528	0.145

Table 7: Power comparison of rare variant set analysis: $n_e = n_c = 2500$, $a_0 = 1, b_0 = 25$, unequal proportions of protective and deleterious variants. The highest powered tests in each row are bold-faced.

$\theta = 0.05, d = -0.6$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.086	0.07	**0.105**	0.084	0.003
10^{-5}	0.129	0.111	**0.151**	0.129	0.008
10^{-4}	0.186	0.166	**0.21**	0.188	0.016
10^{-3}	0.278	0.249	**0.303**	0.278	0.037
$\theta = 0.1, d = -0.5$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.148	0.126	**0.172**	0.145	0.011
10^{-5}	0.211	0.189	**0.239**	0.213	0.021
10^{-4}	0.306	0.277	**0.337**	0.306	0.04
10^{-3}	0.428	0.399	**0.459**	0.432	0.082
$\theta = 0.2, d = -0.4$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.251	0.232	**0.280**	0.260	0.052
10^{-5}	0.339	0.320	**0.376**	0.353	0.082
10^{-4}	0.459	0.445	**0.494**	0.477	0.131
10^{-3}	0.603	0.596	**0.635**	0.625	0.217
$\theta = 0.5, d = -0.25$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.322	0.38	0.358	**0.408**	0.224
10^{-5}	0.431	0.505	0.468	**0.535**	0.313
10^{-4}	0.566	0.648	0.596	**0.672**	0.426
10^{-3}	0.718	0.788	0.744	**0.806**	0.565

Table 8: Power comparison of rare variant set analysis: $n_e = n_c = 2500$, $a_0 = 0.5, b_0 = 24.5$, unequal proportions of protective and deleterious variants. The highest powered tests in each row are bold-faced.

$\theta = 0.05, d = -0.6$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.015	0.01	**0.024**	0.017	0
10^{-5}	0.027	0.019	**0.042**	0.027	0
10^{-4}	0.055	0.041	**0.08**	0.059	0.002
10^{-3}	0.119	0.098	**0.158**	0.129	0.008
$\theta = 0.1, d = -0.5$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.045	0.03	**0.066**	0.046	0
10^{-5}	0.079	0.057	**0.108**	0.081	0.002
10^{-4}	0.141	0.117	**0.181**	0.15	0.007
10^{-3}	0.26	0.225	**0.312**	0.272	0.026
$\theta = 0.2, d = -0.4$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.125	0.101	**0.164**	0.132	0.008
10^{-5}	0.197	0.167	**0.246**	0.21	0.019
10^{-4}	0.312	0.288	**0.368**	0.339	0.046
10^{-3}	0.481	0.467	**0.54**	0.519	0.108
$\theta = 0.5, d = -0.25$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.195	0.234	0.241	**0.271**	0.092
10^{-5}	0.29	0.359	0.348	**0.404**	0.166
10^{-4}	0.437	0.539	0.489	**0.58**	0.283
10^{-3}	0.622	0.73	0.671	**0.762**	0.459

Table 9: Power comparison of rare variant set analysis: $n_e = n_c = 2500$, $a_0 = 1.5, b_0 = 25.5$, unequal proportions of protective and deleterious variants. The highest powered tests in each row are bold-faced.

$\theta = 0.05, d = -0.6$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.074	0.066	**0.087**	0.079	0.013
10^{-5}	0.108	0.1	**0.121**	0.112	0.021
10^{-4}	0.155	0.142	**0.167**	0.156	0.038
10^{-3}	0.221	0.207	**0.234**	0.22	0.07
$\theta = 0.1, d = -0.5$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.116	0.113	**0.132**	0.125	0.03
10^{-5}	0.164	0.158	**0.178**	0.172	0.046
10^{-4}	0.233	0.222	**0.25**	0.239	0.074
10^{-3}	0.328	0.316	**0.346**	0.334	0.126
$\theta = 0.2, d = -0.4$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.18	0.186	0.197	**0.204**	0.086
10^{-5}	0.242	0.248	0.263	**0.267**	0.118
10^{-4}	0.336	0.338	0.354	**0.359**	0.167
10^{-3}	0.461	0.461	0.479	**0.48**	0.243
$\theta = 0.5, d = -0.25$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.225	0.292	0.245	**0.313**	0.229
10^{-5}	0.302	0.382	0.324	**0.403**	0.298
10^{-4}	0.408	0.497	0.431	**0.516**	0.386
10^{-3}	0.554	0.637	0.573	**0.655**	0.505

Table 10: Power comparison of rare variant set analysis: $n_e = 1700, n_c = 3300$, $a_0 = 1, b_0 = 25$, equal proportions of protective and deleterious variants. The highest powered tests in each row are bold-faced.

$\theta = 0.05, d = -0.6$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.162	0.136	**0.166**	0.136	0.005
10^{-5}	0.217	0.191	**0.224**	0.194	0.01
10^{-4}	0.293	0.264	**0.305**	0.273	0.023
10^{-3}	0.403	0.373	**0.419**	0.385	0.053
$\theta = 0.1, d = -0.5$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.236	0.2	**0.246**	0.204	0.012
10^{-5}	0.324	0.289	**0.334**	0.294	0.023
10^{-4}	0.432	0.399	**0.447**	0.409	0.049
10^{-3}	0.57	0.536	**0.584**	0.549	0.091
$\theta = 0.2, d = -0.4$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.31	0.263	**0.325**	0.267	0.021
10^{-5}	0.425	0.382	**0.441**	0.39	0.038
10^{-4}	0.564	0.524	**0.584**	0.537	0.07
10^{-3}	0.718	0.683	**0.734**	0.695	0.128
$\theta = 0.5, d = -0.25$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.264	0.219	**0.279**	0.225	0.017
10^{-5}	0.384	0.338	**0.402**	0.35	0.034
10^{-4}	0.536	0.49	**0.556**	0.509	0.065
10^{-3}	0.713	0.676	**0.73**	0.691	0.124

Table 11: Power comparison of rare variant set analysis: $n_e = 1700, n_c = 3300$, $a_0 = 0.5, b_0 = 24.5$, equal proportions of protective and deleterious variants. The highest powered tests in each row are bold-faced.

$\theta = 0.05, d = -0.6$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.082	0.052	**0.086**	0.052	0.001
10^{-5}	0.141	0.099	**0.147**	0.103	0.001
10^{-4}	0.229	0.186	**0.24**	0.192	0.005
10^{-3}	0.373	0.319	**0.393**	0.336	0.016
$\theta = 0.1, d = -0.5$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.171	0.122	**0.183**	0.125	0.002
10^{-5}	0.268	0.206	**0.287**	0.212	0.005
10^{-4}	0.405	0.341	**0.426**	0.357	0.014
10^{-3}	0.584	0.52	**0.603**	0.535	0.039
$\theta = 0.2, d = -0.4$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.27	0.2	**0.294**	0.209	0.004
10^{-5}	0.408	0.321	**0.43**	0.339	0.01
10^{-4}	0.567	0.494	**0.589**	0.514	0.024
10^{-3}	0.742	0.688	**0.764**	0.705	0.066
$\theta = 0.5, d = -0.25$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.216	0.157	**0.238**	0.164	0.004
10^{-5}	0.342	0.267	**0.369**	0.281	0.01
10^{-4}	0.512	0.44	**0.54**	0.463	0.024
10^{-3}	0.704	0.644	**0.728**	0.668	0.064

Table 12: Power comparison of rare variant set analysis: $n_e = 1700, n_c = 3300$, $a_0 = 1.5, b_0 = 25.5$, equal proportions of protective and deleterious variants. The highest powered tests in each row are bold-faced.

$\theta = 0.05, d = -0.6$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.106	0.1	**0.109**	0.1	0.016
10^{-5}	0.142	0.132	**0.145**	0.133	0.027
10^{-4}	0.186	0.175	**0.192**	0.179	0.045
10^{-3}	0.257	0.243	**0.264**	0.25	0.081
$\theta = 0.1, d = -0.5$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.144	0.134	**0.145**	0.132	0.026
10^{-5}	0.195	0.183	**0.203**	0.188	0.043
10^{-4}	0.27	0.255	**0.277**	0.261	0.068
10^{-3}	0.369	0.353	**0.381**	0.36	0.115
$\theta = 0.2, d = -0.4$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.162	0.153	**0.17**	0.159	0.035
10^{-5}	0.237	0.222	**0.246**	0.228	0.057
10^{-4}	0.338	0.325	**0.352**	0.333	0.09
10^{-3}	0.476	0.456	**0.488**	0.466	0.149
$\theta = 0.5, d = -0.25$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.137	0.13	**0.143**	0.132	0.031
10^{-5}	0.204	0.192	**0.214**	0.199	0.05
10^{-4}	0.312	0.295	**0.325**	0.302	0.084
10^{-3}	0.473	0.451	**0.488**	0.458	0.142

Table 13: Power comparison of rare variant set analysis: $n_e = 1700, n_c = 3300$, $a_0 = 1, b_0 = 25$, unequal proportions of protective and deleterious variants. The highest powered tests in each row are bold-faced.

$\theta = 0.05, d = -0.6$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.04	0.031	**0.063**	0.048	0.001
10^{-5}	0.071	0.056	**0.1**	0.084	0.002
10^{-4}	0.118	0.101	**0.155**	0.136	0.005
10^{-3}	0.196	0.174	**0.237**	0.215	0.018
$\theta = 0.1, d = -0.5$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.075	0.06	**0.109**	0.088	0.004
10^{-5}	0.124	0.103	**0.169**	0.147	0.008
10^{-4}	0.201	0.178	**0.254**	0.231	0.018
10^{-3}	0.318	0.291	**0.378**	0.351	0.047
$\theta = 0.2, d = -0.4$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.129	0.112	**0.186**	0.168	0.017
10^{-5}	0.207	0.192	**0.276**	0.263	0.036
10^{-4}	0.318	0.31	**0.391**	0.382	0.066
10^{-3}	0.473	0.464	**0.542**	0.537	0.136
$\theta = 0.5, d = -0.25$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.155	0.21	0.234	**0.296**	0.118
10^{-5}	0.251	0.333	0.337	**0.427**	0.191
10^{-4}	0.382	0.495	0.47	**0.576**	0.303
10^{-3}	0.556	0.671	0.641	**0.736**	0.456

Table 14: Power comparison of rare variant set analysis: $n_e = 1700, n_c = 3300$, $a_0 = 0.5, b_0 = 24.5$, unequal proportions of protective and deleterious variants. The highest powered tests in each row are bold-faced.

$\theta = 0.05, d = -0.6$					
α	SKAT	SKAT-O	SKAT$_L$	SKAT-O$_L$	Burden
10^{-6}	0.006	0.003	**0.01**	0.005	0
10^{-5}	0.012	0.008	**0.021**	0.013	0
10^{-4}	0.03	0.02	**0.046**	0.033	0.001
10^{-3}	0.074	0.056	**0.108**	0.084	0.004
$\theta = 0.1, d = -0.5$					
α	SKAT	SKAT-O	SKAT$_L$	SKAT-O$_L$	Burden
10^{-6}	0.012	0.007	**0.026**	0.014	0
10^{-5}	0.028	0.015	**0.051**	0.035	0.001
10^{-4}	0.062	0.044	**0.105**	0.078	0.002
10^{-3}	0.15	0.121	**0.209**	0.18	0.014
$\theta = 0.2, d = -0.4$					
α	SKAT	SKAT-O	SKAT$_L$	SKAT-O$_L$	Burden
10^{-6}	0.034	0.023	**0.069**	0.052	0.001
10^{-5}	0.069	0.051	**0.124**	0.1	0.006
10^{-4}	0.143	0.124	**0.216**	0.199	0.02
10^{-3}	0.282	0.272	**0.375**	0.371	0.059
$\theta = 0.5, d = -0.25$					
α	SKAT	SKAT-O	SKAT$_L$	SKAT-O$_L$	Burden
10^{-6}	0.046	0.065	0.099	**0.129**	0.036
10^{-5}	0.092	0.143	0.166	**0.233**	0.08
10^{-4}	0.176	0.293	0.277	**0.406**	0.174
10^{-3}	0.334	0.522	0.448	**0.624**	0.333

Table 15: Power comparison of rare variant set analysis: $n_e = 1700, n_c = 3300$, $a_0 = 1.5, b_0 = 25.5$, unequal proportions of protective and deleterious variants. The highest powered tests in each row are bold-faced.

$\theta = 0.05, d = -0.6$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.035	0.032	**0.051**	0.046	0.004
10^{-5}	0.057	0.052	**0.08**	0.072	0.006
10^{-4}	0.096	0.087	**0.121**	0.112	0.017
10^{-3}	0.152	0.143	**0.177**	0.17	0.039
$\theta = 0.1, d = -0.5$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.054	0.051	**0.077**	0.073	0.011
10^{-5}	0.088	0.083	**0.119**	0.114	0.019
10^{-4}	0.149	0.14	**0.184**	0.178	0.04
10^{-3}	0.238	0.23	**0.277**	0.27	0.079
$\theta = 0.2, d = -0.4$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.087	0.091	0.121	**0.13**	0.034
10^{-5}	0.139	0.145	0.18	**0.192**	0.057
10^{-4}	0.223	0.231	0.272	**0.284**	0.099
10^{-3}	0.353	0.358	0.399	**0.408**	0.167
$\theta = 0.5, d = -0.25$					
α	SKAT	SKAT-O	SKAT_L	SKAT-O_L	Burden
10^{-6}	0.102	0.163	0.147	**0.22**	0.124
10^{-5}	0.164	0.241	0.215	**0.308**	0.185
10^{-4}	0.26	0.362	0.319	**0.432**	0.278
10^{-3}	0.406	0.521	0.465	**0.579**	0.402

2 Significance p-value calculation

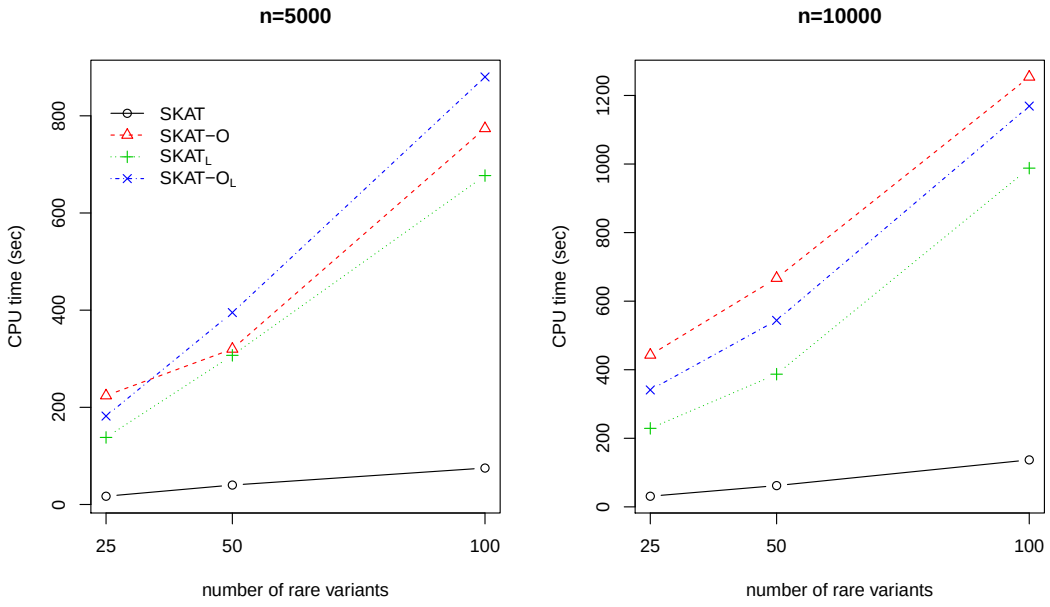
For the proposed SKAT$_L$, the p-values can be computed based on the mixture of 1-DF chi-square distributions. We use the Davies' method (Davies, 1980) implemented in the 'CompQuadForm' R package (Lafaye De Micheaux, 2013). When the Davies' method failed to converge, we use the saddlepoint approximation method (Kuonen, 1999) implemented in the 'survey' R package (Lumley, 2014) following the approach of Chen *et al.* (2014). When computing p-values for the proposed SKAT-O$_L$, we use the chi-square approximation to compute the quantile for the mixture of 1-DF chi-square distributions following the approach of Lee *et al.* (2012). The one-dimensional integration is then numerically computed based on the convolution of SKAT$_L$ p-value and 1-DF chi-square distribution.

3 Numerical computation

Figure 2 shows the average CPU sec used to compute significance p-values for 1000 rare variant sets on a single Linux workstation with 3 GHz CPU and 8 GB memory. We follow the previous simulation setup with 2 covariates. We consider two sample sizes: $n = 5000$ and $n = 10000$ samples; and three variant set sizes: 25, 50 and 100 variants in each variant set. We repeated ten times and reported the average computing time. Overall we can see that all methods roughly scale linearly with the variant set size. The SKAT is the most efficient, and the proposed methods have comparable speed as the SKAT-O approach.

For the ARIC diabetes data, when analyzing 1415 variant sets on chromosome 1 on the same Linux workstation, SKAT takes 42 sec CPU time, SKAT-O takes 674 sec CPU time, SKAT$_L$ takes 151 sec CPU time, and SKAT-O$_L$ takes 630 sec CPU time on the same machine. In our simulation study, on average around 50% of the variants in a set have their minor allele counts less than 10 and we use their standardized score statistics to replace their likelihood ratio test statistics (LRT). When using the LRT for all rare variants in the proposed methods, the type I errors are significantly inflated (ratio of estimated type I errors over nominal significance level α could be as high as 3.0).

Figure 2: Average CPU sec used to analyze 1000 variant sets: the x-axis shows the variant set size, the y-axis shows the average CPU sec used (over 10 simulations); the left plot is for 5000 samples and 2 covariates following our simulation setup, the right plot is for 10000 samples.



4 ARIC data analysis

One reviewer raised the concern that our diabetes analysis did not adjust for BMI, and the identified association between ZZZ3 and type 2 diabetes could be confounded by BMI. When analyzing the ARIC type 2 diabetes outcome, we followed the convention of adjusting for the age, gender and field center (for potential population stratification/ethnicity). Initial GWAS analyses of type 2 diabetes in the ARIC white population also included adjustment principal components of ancestry in addition to the field center, but these PCA adjustments were found to make no difference in the effect estimates and were therefore not routinely included in subsequent analyses. The BMI is not included as a covariate for adjustment. In the DIAGRAM (DIAbetes Genetics Replication And Meta-analysis) consortium, the largest international initiative to characterize the genetic basis of type 2 diabetes, the vast majority of studies (including the ARIC Study) did not adjust for BMI or any other measure of adiposity in efforts that led to discovery of several dozen new susceptibility loci for type 2 diabetes (Voight *et al.*, 2010; Morris

Table 16: Type I error of rare variant set analysis under 1:6 and 1:10 case-control ratio. The Type I errors have been divided by their nominal levels.

case:control	1:6			1:10		
α	10^{-5}	10^{-4}	10^{-3}	10^{-5}	10^{-4}	10^{-3}
SKAT	3.20	1.78	1.28	3.60	2.06	1.39
SKAT_L	1.00	1.08	1.06	1.50	1.35	1.33

et al., 2012). Of the 38 studies included in the 2012 DIAGRAM paper, only 4 studies adjusted for BMI and with most performing minimal adjustment for age, sex, and, in some cases, study specific covariates such as field center. It is true that at least one established diabetes locus (FTO) is also an established locus for BMI, but the consensus is that FTO's downstream association with diabetes is mediated by its more proximal effect on BMI. Any risk factor such as BMI that falls on a causal pathway linking the gene variant and type 2 diabetes should not be treated as a confounder for analysis purposes. In the analysis, on average around 60% of the variants in a set have their minor allele counts less than 10 and we use their standardized score statistics to replace their likelihood ratio test statistics.

5 Unbalanced study

As reviewers suggested, we investigate the performance of the proposed methods under very unbalanced case-control ratios, 1:6 (mimicking the ARIC diabetes data) and 1:10. We use Beta(1,25) weight for SKAT and Beta(1.5,25.5) weight for SKAT_L. We simulated 10^4 individuals and performed 10^7 simulations to estimate the type I errors. Table 16 summarizes the results. Under 1:6 case-control ratio, the proposed SKAT_L has better controlled type I errors than SKAT, which has increased inflations under smaller significance level. Under 1:10 case-control ratio, both methods have inflated type I errors.

We performed 10^4 simulations to estimate the power under 1:6 case-control ratio with equal proportions of protective and deleterious variants. Table 7 summarizes the results. Overall we can see that the proposed SKAT_L has larger power than SKAT.

Overall the proposed SKAT_L has the best performance under balanced case-control study,

Table 17: Power comparison of rare variant set analysis under 1:6 case-control ratio.

	$\theta = 0.05, d = -0.6$		$\theta = 0.1, d = -0.5$		$\theta = 0.2, d = -0.4$		$\theta = 0.5, d = -0.25$	
α	SKAT	SKAT_L	SKAT	SKAT_L	SKAT	SKAT_L	SKAT	SKAT_L
10^{-6}	0.372	0.395	0.529	0.561	0.660	0.697	0.657	0.699
10^{-5}	0.414	0.473	0.580	0.651	0.712	0.780	0.711	0.784
10^{-4}	0.541	0.566	0.722	0.748	0.844	0.864	0.848	0.869

and both SKAT and SKAT_L have inflated type I errors under highly unbalanced case-control ratios, which are consistent with the performance of score test and likelihood ratio test in single rare variant association test (Ma *et al.*, 2013). More research is needed to develop methods that have appropriately calibrated p-values under unbalanced case-control ratios. Exactly computing the analytical p-value of SKAT-O is in general very difficult, and we have to employ some analytical and numerical approximations. Our simulation results have shown that the proposed SKATO_L could control the type I errors though with conservative performance. SKAT-O was shown to have slightly inflated type I errors (Lee *et al.*, 2012, Table 2, p. 769). More research is needed to develop computational methods that can quickly compute better calibrated significance p-values for SKAT-O.

References

Chen,H., Lumley,T., Brody,J., Heard-Costa,N.L., Fox,C.S., Cupples,L.A. and Dupuis,J. (2014) Sequence kernel association test for survival traits. *Genetic Epidemiology,* **38** (3), 191–197.

Davies,R.B. (1980) Algorithm AS 155: the distribution of a linear combination of χ^2 random variables. *Applied Statistics,* **29** (3), 323.

Kuonen, D., 1999. Saddlepoint approximations for distributions of quadratic forms in normal variables. Biometrika 86 (4), 929–935.

Lafaye De Micheaux, P., 2013. CompQuadForm: Distribution function of quadratic forms in normal variables. R package version 1.4.1. `http://cran.r-project.org/web/packages/CompQuadForm/index.html`

Lee,S., Wu,M.C. and Lin,X. (2012) Optimal tests for rare variant effects in sequencing association studies. *Biostatistics,* **13** (4), 762–775.

Lumley, T., 2014. survey: analysis of complex survey samples. R package version 3.30-3. `http://cran.r-project.org/web/packages/survey/index.html`

Ma,C., Blackwell,T., Boehnke,M., Scott,L.J. and GoT2D Investigators. (2013) Recommended joint and meta-analysis strategies for case-control association testing of single low-count variants. *Genetic Epidemiology,* **37** (6), 539–550.

Morris,A.P., Voight,B.F., Teslovich,T.M., Ferreira,T., Segre,A.V. and others. (2012) Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature Genetics,* **44** (9), 981–990.

Schaffner,S.F., Foo,C., Gabriel,S., Reich,D., Daly,M.J. and Altshuler,D. (2005) Calibrating a coalescent simulation of human genome sequence variation. *Genome Research,* **15** (11), 1576–1583.

Voight,B.F., Scott,L.J., Steinthorsdottir,V., Morris,A.P., Dina,C., Welch,R.P., and others. (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature Genetics,* **42** (7), 579–589.

Wu,M., Lee,S., Cai,T., Li,Y., Boehnke,M. and Lin,X. (2011) Rare-variant association testing for sequencing data with the sequence kernel association test. *The American Journal of Human Genetics,* **89** (1), 82–93.