

Supplementary Tables

Article Title: Estimating causal effects of rare treatments on binary outcomes: Addressing sample size requirements and bias correction

Table S1: Estimates of IPW-ATE over varying prevalence of treatment variable (p) and EPV for a fixed sample size ($n = 250, 500, 1000$) and fixed number of baseline covariates ($k = 3$). True ATE = 0.443. The results are summarized over 10000 simulations.

n	Exp% (EPV)	Correctly Specified Model				Misspecified model				Overfitted model			
		bias	robSE	simSE	ratio	bias	robSE	simSE	ratio	bias	robSE	simSE	ratio
250	3 (2)	-0.028	0.992	1.454	0.682	-0.142	0.965	1.379	0.700	-0.041	0.998	1.515	0.659
	5 (3.1)	0.016	0.834	1.140	0.732	-0.111	0.816	1.121	0.728	0.019	0.845	1.171	0.722
	8 (5)	-0.016	0.662	0.812	0.815	-0.109	0.644	0.798	0.807	-0.010	0.671	0.836	0.803
	10 (6.3)	0.006	0.586	0.657	0.892	-0.103	0.575	0.654	0.879	-0.005	0.593	0.676	0.877
	12 (7.5)	0.000	0.532	0.578	0.920	-0.094	0.522	0.554	0.942	-0.005	0.538	0.602	0.894
	15 (9.4)	-0.005	0.474	0.495	0.958	-0.073	0.465	0.501	0.928	-0.003	0.480	0.505	0.950
	18 (11.3)	0.006	0.425	0.426	0.998	-0.056	0.417	0.412	1.012	0.005	0.428	0.430	0.995
	20 (12.4)	-0.005	0.411	0.413	0.995	-0.071	0.401	0.395	1.015	-0.006	0.415	0.419	0.990
500	3 (3.8)	-0.027	0.774	1.001	0.773	-0.140	0.763	1.059	0.720	-0.023	0.784	1.040	0.754
	5 (6.2)	0.004	0.605	0.688	0.879	-0.119	0.601	0.698	0.861	0.011	0.611	0.704	0.868
	8 (10.1)	-0.009	0.465	0.487	0.955	-0.118	0.464	0.495	0.937	-0.014	0.471	0.496	0.950
	10 (12.4)	-0.014	0.411	0.438	0.938	-0.128	0.416	0.436	0.954	-0.015	0.414	0.443	0.935
	12 (15)	0.008	0.372	0.389	0.956	-0.091	0.370	0.378	0.979	0.008	0.375	0.389	0.964
	15 (18.8)	0.004	0.328	0.326	1.006	-0.067	0.326	0.324	1.006	0.005	0.330	0.330	1.000
	18 (22.6)	0.009	0.297	0.288	1.031	-0.056	0.292	0.279	1.047	0.010	0.299	0.290	1.031
	20 (24.8)	0.006	0.285	0.275	1.036	-0.056	0.280	0.268	1.045	0.005	0.286	0.275	1.040
1000	3 (7.5)	0.007	0.558	0.636	0.877	-0.148	0.567	0.702	0.808	0.005	0.563	0.649	0.867
	5 (12.4)	-0.003	0.421	0.430	0.979	-0.176	0.437	0.483	0.905	-0.003	0.424	0.437	0.970
	8 (20.1)	-0.009	0.321	0.314	1.022	-0.144	0.331	0.335	0.988	-0.009	0.324	0.318	1.019
	10 (25)	0.002	0.286	0.279	1.025	-0.115	0.294	0.296	0.993	0.001	0.287	0.281	1.021
	12 (30)	0.008	0.259	0.261	0.992	-0.099	0.262	0.259	1.012	0.008	0.260	0.262	0.992
	15 (37.4)	0.012	0.229	0.224	1.022	-0.079	0.231	0.228	1.013	0.012	0.230	0.225	1.022
	18 (45.2)	0.002	0.208	0.202	1.030	-0.071	0.207	0.197	1.051	0.002	0.209	0.203	1.030
	20 (49.5)	0.001	0.199	0.188	1.059	-0.052	0.196	0.179	1.095	0.002	0.200	0.188	1.064

$ratio = robSE / simSE$

Table S2: Estimates of IPW-ATE over varying prevalence of treatment variable (p) and EPV for a fixed sample size ($n = 250, 500, 1000$) and fixed number of baseline covariates ($k = 6$). True ATE = 0.386. The results are summarized over 10000 simulations.

n	$p(\%)$	EPV	Correctly specified Model				Misspecified Model				Overfitted Model			
			bias	simSE	robSE	ratio	bias	simSE	robSE	ratio	bias	simSE	robSE	ratio
250	2	1.1	0.097	2.030	1.090	0.537	0.166	1.781	1.065	0.598	0.099	2.077	1.091	0.537
	3	1.4	0.097	1.893	1.042	0.550	0.182	1.632	1.010	0.619	0.093	1.926	1.045	0.550
	5	2.1	0.083	1.596	0.937	0.587	0.163	1.410	0.895	0.635	0.085	1.634	0.941	0.587
	8	3.4	0.078	1.161	0.788	0.679	0.090	1.065	0.744	0.699	0.075	1.196	0.793	0.679
	10	4.1	0.067	0.951	0.718	0.754	0.106	0.835	0.672	0.804	0.076	0.985	0.723	0.754
	12	5.0	0.053	0.807	0.659	0.816	0.100	0.711	0.618	0.869	0.074	0.845	0.666	0.816
	15	6.2	0.052	0.687	0.585	0.852	0.078	0.619	0.542	0.875	0.067	0.712	0.592	0.852
	18	7.5	0.052	0.602	0.537	0.892	0.088	0.524	0.508	0.969	0.049	0.618	0.541	0.892
	20	8.4	0.042	0.553	0.507	0.917	0.079	0.498	0.471	0.946	0.059	0.556	0.511	0.917
500	2	1.7	0.058	1.790	1.009	0.564	0.190	1.608	0.966	0.601	0.059	1.841	1.014	0.564
	3	2.6	0.044	1.480	0.896	0.605	0.169	1.290	0.857	0.664	0.047	1.545	0.901	0.605
	5	4.2	0.042	1.056	0.737	0.698	0.142	0.997	0.701	0.703	0.050	1.080	0.743	0.698
	8	6.7	0.041	0.784	0.588	0.750	0.106	0.700	0.560	0.800	0.049	0.794	0.594	0.750
	10	8.3	0.040	0.604	0.532	0.880	0.107	0.568	0.503	0.885	0.048	0.610	0.535	0.880
	12	10.0	0.020	0.537	0.480	0.893	0.085	0.501	0.454	0.905	0.020	0.546	0.482	0.893
	15	12.5	0.032	0.481	0.426	0.885	0.070	0.435	0.400	0.919	0.031	0.485	0.428	0.885
	18	15.0	-0.006	0.396	0.387	0.978	0.054	0.339	0.360	1.062	-0.007	0.401	0.389	0.978
	20	16.6	-0.017	0.361	0.363	1.005	0.035	0.328	0.341	1.040	-0.017	0.362	0.365	1.005
1000	2	3.4	0.057	1.282	0.820	0.640	0.108	1.245	0.783	0.629	0.064	1.317	0.826	0.640
	3	5.0	0.049	0.912	0.689	0.755	0.145	0.935	0.670	0.716	0.060	0.922	0.695	0.755
	5	8.3	0.048	0.693	0.560	0.809	0.131	0.646	0.537	0.833	0.058	0.698	0.563	0.809
	8	13.3	0.047	0.464	0.430	0.927	0.116	0.445	0.417	0.939	0.051	0.467	0.432	0.927
	10	16.7	0.042	0.419	0.385	0.917	0.112	0.393	0.368	0.937	0.013	0.422	0.386	0.917
	12	20.0	0.022	0.372	0.345	0.927	0.081	0.355	0.330	0.930	0.008	0.376	0.346	0.927
	15	25.0	0.021	0.308	0.305	0.991	0.084	0.285	0.289	1.016	0.021	0.308	0.306	0.991
	18	30.1	0.014	0.262	0.271	1.037	0.063	0.253	0.258	1.020	0.025	0.263	0.272	1.037
	20	33.4	0.016	0.246	0.259	1.053	0.063	0.223	0.244	1.091	0.026	0.246	0.259	1.053

$ratio = robSE / simSE$

Table S3: Standardized Mean Difference (SMD) for checking covariate balancing between LBW and normal birth weight while estimating IPSW-ATE for LBW on stunting.

Covariates	Categories	Balance before weighting			Balance after weighting		
		Normal	LBW	SMD	Normal	LBW	SMD
	<i>n</i>	1765	318		2084	2065	
Mother's Education	no education	47 (2.7)	18 (5.8)	0.212	66 (3.2)	67 (3.3)	0.034
	primary	270 (15.3)	61 (19.2)		331 (15.9)	334 (16.2)	
	secondary	928 (52.6)	163 (51.2)		1092 (52.4)	1105 (53.5)	
	higher	520 (29.5)	76 (23.8)		595 (28.5)	558 (27.0)	
Mother's Height	≤ 145 cm	208 (10.59)	68(18.4)	0.05	264 (11.78)	265 (11.83)	0.009
	> 145 cm	1754 (89.41)	310 (81.96)		1974 (88.22)	1975 (88.17)	
Antenatal Care	<4	592 (33.5)	124 (38.9)	0.112	716.0 (34.3)	694.9 (33.6)	0.015
	≥4	1174 (66.5)	195 (61.1)		1369 (65.7)	1371 (66.4)	
Socio economic status	poorest	186 (10.5)	50 (15.8)	0.206	236 (11.3)	245 (11.9)	0.03
	poorer	261 (14.8)	49 (15.3)		310 (14.9)	314 (15.2)	
	middle	337 (19.1)	66 (20.6)		403 (19.4)	411 (19.9)	
	richer	434 (24.6)	79 (24.8)		513 (24.6)	490 (23.7)	
	richest	547 (31.0)	75 (23.6)		622 (29.8)	606 (29.3)	
Division	Barisal	83 (4.7)	13 (4.2)	0.215	96 (4.6)	95 (4.6)	0.043
	Chittagong	292 (16.6)	73 (23.0)		366 (17.6)	373 (18.1)	
	Dhaka	527 (29.8)	90 (28.2)		616.1 (29.5)	593.1 (28.7)	
	Khulna	211 (11.9)	37 (11.6)		248.1 (11.9)	253.5 (12.3)	
	Mymenshing	147 (8.3)	16 (5.0)		163 (7.8)	164 (8.0)	
	Rajshahi	203 (11.5)	40 (12.4)		242.2 (11.6)	235.6 (11.4)	
	Rangpur	230 (13.0)	34 (10.6)		263 (12.6)	248 (12.0)	
	Sylhet	73 (4.2)	15 (4.9)		89 (4.3)	102 (5.0)	