

Covariate Use Across All Models

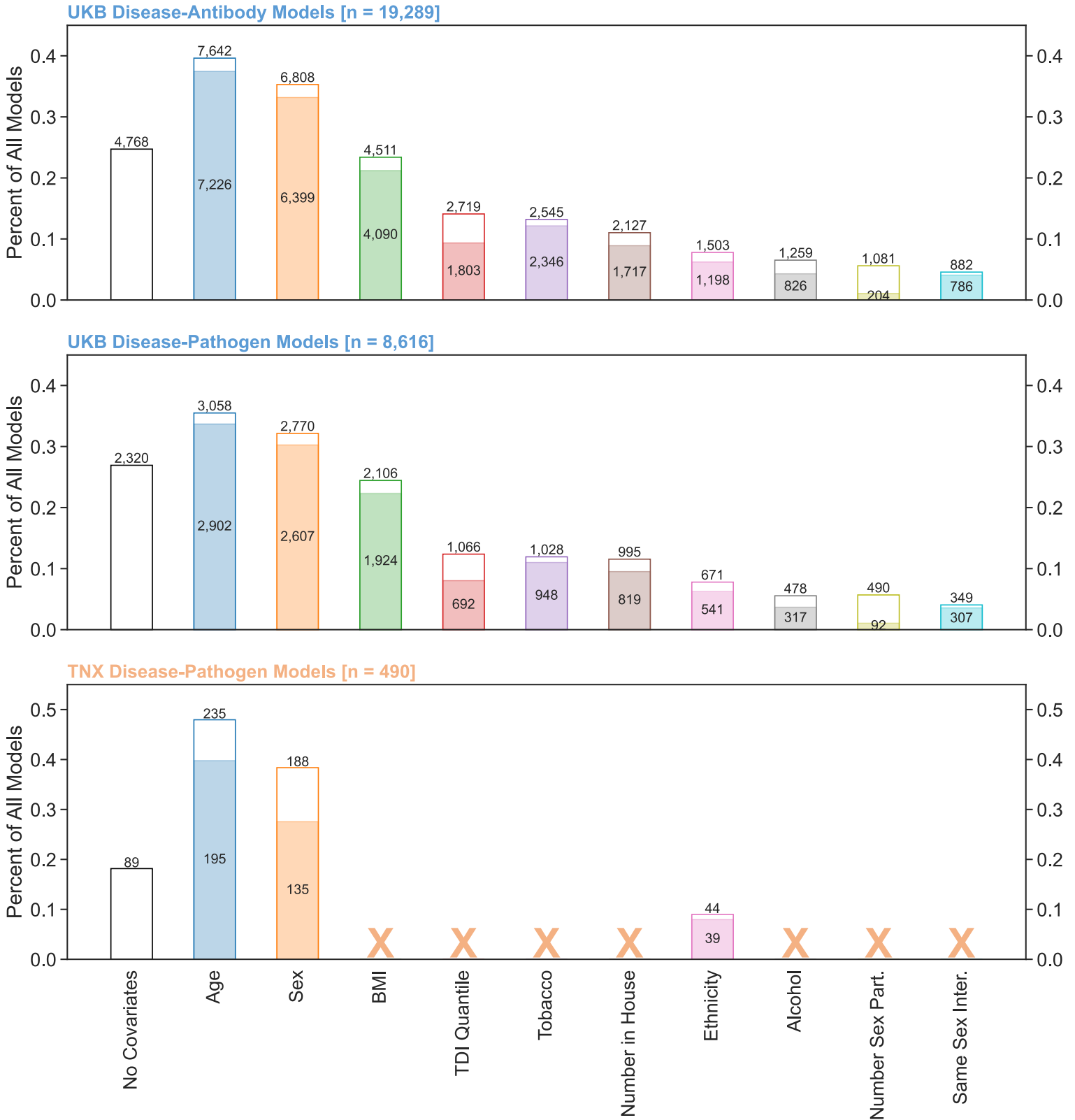


Figure S1 | Covariate use across models.

Bar charts summarizing covariate usage in antibody-disease and pathogen-disease tests across the discovery and replication cohorts. The height of each bar indicates the percentage of all models that were adjusted for the corresponding sociodemographic or health-related covariate. In order to be included in an antibody-disease model for adjustment, a covariate had to be significantly associated (unadjusted $p < 0.05$) in univariate tests with both the disease status and the titer level separately (Supplemental Table 2). The filled portion of each bar indicates the percent of the final multivariate logistic regression models in which the covariate remained significant (unadjusted $p < 0.05$). The covariates that were unavailable in the TNX cohort are marked by an orange "X".

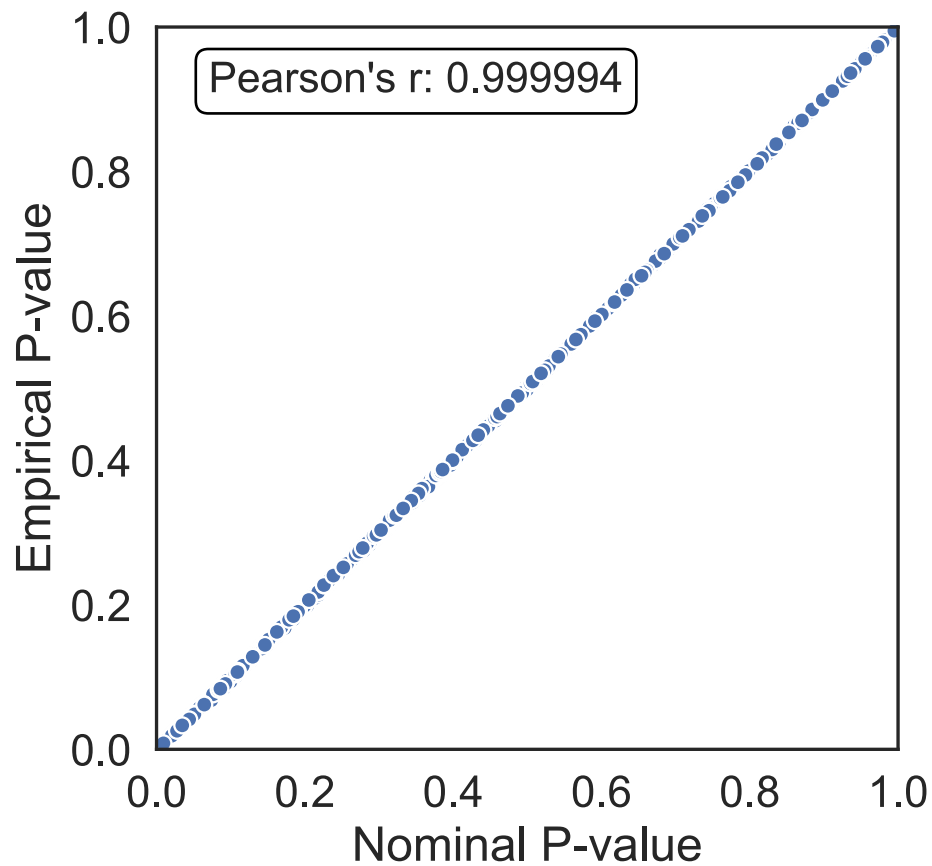


Figure S2 | Nominal versus empirical p-values across all UK Biobank antibody-disease models.

Comparison of the nominal and empirical p-values for all discovery cohort antibody-disease models. To calculate empirical p-values, 10,000 permutations of each antibody-disease model were performed combining all permutations for a particular disease into a per-disease null distribution, yielding 450,000 permutation results (see Methods). The nominal p-value for a specific antibody-disease model was compared to the disease-specific null distribution to calculate the empirical p-value.

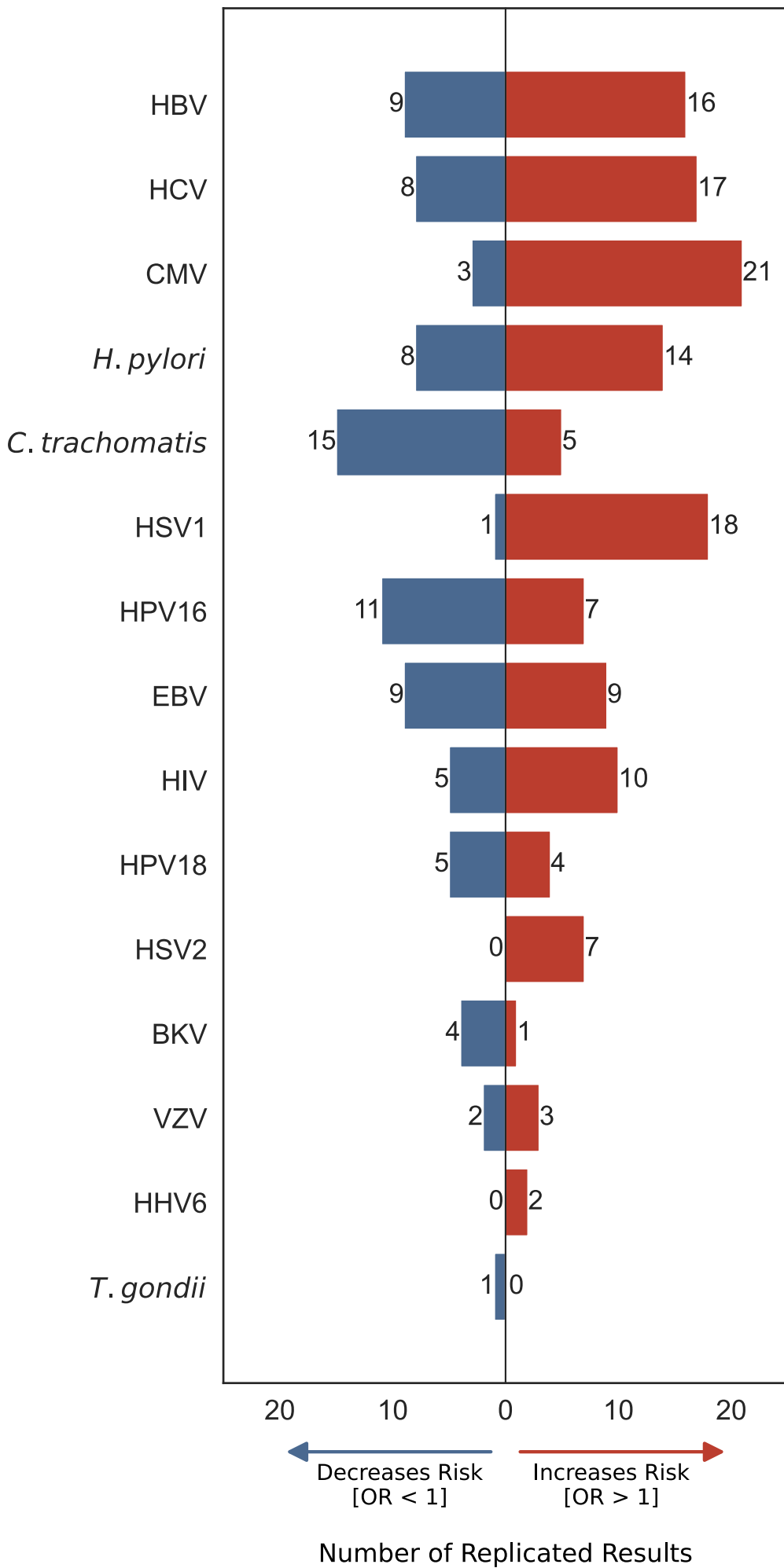


Figure S3 | Replicated results split by pathogen and effect direction.

Split bar chart showing the number of replicated “risk” and “protective” associations found for each of the pathogens. The total number of replicated results for each pathogen are split by effect direction. Those with an odds ratio less than one are represented by the blue bars to the left of the center line. Those with an odds ratio greater than one are represented by the red bars to the right of the center line.

Replicated Results
[Protective | Risk]

HCV	1 [0 1]	2 [2 0]	1 [0 1]	1 [0 1]	1 [0 1]	2 [1 1]		1 [0 1]	3 [1 2]		7 [3 4]	1 [0 1]	2 [1 1]	3 [0 3]	
EBV	1 [0 1]	1 [1 0]		2 [2 0]		1 [0 1]		1 [1 0]	1 [1 0]	3 [1 2]	1 [1 0]	1 [1 0]	3 [0 3]	3 [1 2]	
HBV	1 [0 1]					3 [1 2]		1 [0 1]	4 [3 1]	1 [0 1]	5 [3 2]	3 [1 2]	3 [1 2]	3 [0 3]	1 [0 1]
HPV16						1 [1 0]	1 [1 0]	1 [0 1]	1 [1 0]	2 [1 1]	4 [3 1]	2 [1 1]	4 [3 1]	2 [0 2]	
HSV1	1 [0 1]		1 [0 1]	1 [0 1]		1 [1 0]				2 [0 2]	1 [0 1]	3 [0 3]	5 [0 5]	4 [0 4]	
<i>C. trachomatis</i>		1 [1 0]				1 [1 0]			2 [2 0]	3 [3 0]	3 [3 0]	2 [2 0]	4 [3 1]	4 [0 4]	
<i>H. pylori</i>			2 [0 2]	1 [0 1]		1 [1 0]			2 [1 1]	2 [2 0]	11 [4 7]		1 [0 1]	2 [0 2]	
HIV	1 [0 1]						1 [0 1]		1 [1 0]	1 [0 1]	3 [2 1]	2 [0 2]	4 [2 2]	2 [0 2]	
HPV18						1 [1 0]		1 [0 1]	1 [1 0]	1 [1 0]	1 [1 0]	1 [0 1]	1 [1 0]	2 [0 2]	
CMV						2 [2 0]			1 [0 1]	4 [0 4]	8 [1 7]	3 [0 3]	3 [0 3]	3 [0 3]	
VZV	1 [0 1]		1 [0 1]						1 [0 1]			1 [1 0]	1 [1 0]		
BKV										1 [1 0]	1 [1 0]		1 [1 0]	2 [1 1]	
HSV2	2 [0 2]								2 [0 2]				3 [0 3]		
HHV6	1 [0 1]							1 [0 1]							
<i>T. gondii</i>										1 [1 0]					

[A00-B99] Infectious

[C00-D49] Neoplasms

[D50-D89] Blood

[E00-E90] Endocrine, Nutritional, Metabolic

[F00-F99] Mental, Behavioral

[G00-G99] Nervous

[H00-H59] Eye

[H60-H95] Ear

[I00-I99] Circulatory

[J00-J99] Respiratory

[K00-K93] Digestive

[L00-L99] Skin, Subcutaneous

[M00-M99] Musculoskeletal

[N00-N99] Genitourinary

[O00-O99] Pregnancy, Childbirth

Figure S4 | Heatmap of replicated results at the ICD10 chapter level.

A heatmap showing the total number of replicated results for each pathogen across all diseases in an ICD10 chapter, which indicates broad disease classes. All cells are annotated with the total number of replicated results followed by, in square brackets, the number of these replicated results with odds ratios less than one, followed by those with odds ratios greater than one. White squares with no annotation indicate no replicated associations were found for that pathogen and any of the tested diseases in that ICD10 chapter.

		UKB	TNX
Sex	Female	5,276	7,410,388
	Male	4,153	3,807,258
	Other	0	5,827
	Total	9,429	11,223,473

Age	Resolution: Mean	67.5	46.3
	UKB: Month SD	8.1	17.3
	TNX: Year Range	[43.8 - 82.6]	[0 - 91]

Race/Ethnicity	White	8,883	4,654,725
	Asian	231	258,886
	Black	140	1,664,492
	Hispanic	-	860,830
	American Indian or Alaska Native	-	36,516
	Native Hawaiian or Other Pacific Islander	-	12,767
	Other	175	3,735,257
	Total	9,429	11,223,473

BMI	Mean	27.3	-
	SD	4.8	-
	Range	[16.2 - 61.3]	-
n Imputed		32	-

Townsend Deprivation Index	Lowest 1/6	1,572	-
	Middle 2/3	6,285	-
	Highest 1/6	1,572	-
	Total	9,429	0
n Imputed		8	-

Number in House			
Alone		1,712	-
	2	4,331	-
	3	1,524	-
	4	1,282	-
>= 5		514	-
Unknown		66	-
Total		9,429	0

Tobacco Use			
Never		5,210	-
Previous		3,218	-
Current		949	-
Unknown		52	-
Total		9,429	0

Alcohol Use			
Never		417	-
Previous		332	-
Current		8,661	-
Unknown		19	-
Total		9,429	

Number of Sex Partners			
0 partners		88	-
1 partner		2,265	-
2-3		1,838	-
4-5		1,255	-
> 5		2,343	-
Unknown		1,640	-
Total		9,429	0

Same-sex Intercourse			
No		8,204	-
Yes		283	-
Unknown		942	-
Total		9,429	0

Table S1 | Summary statistics for additional covariates across both cohorts.

This table contains the summary statistics for each of the ten additional health-related and sociodemographic variables that were considered as possible confounders. For continuous variables, the mean, standard deviation, and range are presented. For categorical variables, counts for each group are included. Note that the resolution for age differed between cohorts, with UKB having a resolution of one month, while TNX only provided a resolution of one year.

Covariate	Ab Titer Assoc Test	Disease Status Assoc Test
Sex	t-test	Chi-squared
Age	linear regression	t-test
BMI	linear regression	t-test
Race	ANOVA	Chi-squared
Townsend Deprivation Index	ANOVA	Chi-squared
Number in House	ANOVA	Chi-squared
Tobacco Use	ANOVA	Chi-squared
Alcohol Use	ANOVA	Chi-squared
Number of Sex Partners	ANOVA	Chi-squared
Same-sex Intercourse	ANOVA	Chi-squared

Table S2 | Univariate tests used to determine confounding.

This table shows the type of univariate test that was performed between each covariate and the antibody titers as well as each covariate and disease status to determine if the covariate was confounding. For a covariate to be considered a confounder it had to be significantly associated (unadjusted $p < 0.05$) separately with both disease status and antibody titer.