

Table S5. Fairness Metrics Performance

<u>Models</u>		<u>Equalized Odds</u>			<u>Independence</u>			<u>Overall Accuracy Equality</u>			<u>Sufficiency</u>		
		Internal Test	External Test #1	External Test #2	Internal Test	External Test #1	External Test #2	Internal Test	External Test #1	External Test #2	Internal Test	External Test #1	External Test #2
Sex													
	Not Using Sensitive Attributes (M1)	0.273	0.102	0.313	0.134	0.063	0.175	0.085	0.046	0.062	0.101	0.032	0.05
	Using Sensitive Attributes (M2)	0.058	0.429	0.033	0.038	0.176	0.031	0.031	0.093	0.001	0.042	0.132	0.114
	Separate Models for Each Group (M3)	0.358	0.496	0.468	0.173	0.258	0.251	0.107	0.172	0.111	0.103	0.131	0.024
Race													
	Not Using Sensitive Attributes (M1)	0.323	0.277	0.106	0.167	0.117	0.042	0.101	0.068	0.018	0.058	0.102	0.074
	Using Sensitive Attributes (M2)	0.217	0.330	0.049	0.123	0.126	0.012	0.078	0.060	0.002	0.029	0.098	0.063
	Separate Models for Each Group (M3)	0.157	0.500	0.082	0.110	0.257	0.045	0.082	0.172	0.007	0.041	0.110	0.037
Ethnicity													
	Not Using Sensitive Attributes (M1)	0.474	0.57	0.549	0.183	0.111	0.31	0.052	0.043	0.093	0.07	0.246	0.087
	Using Sensitive Attributes (M2)	0.369	0.574	0.642	0.152	0.092	0.351	0.052	0.013	0.089	0.047	0.223	0.146
	Separate Models for Each Group (M3)	0.348	0.357	0.183	0.181	0.174	0.122	0.112	0.127	0.015	0.065	0.078	0.184