

Table S4a. Performance Metrics

Models		AUC			ACC			Precision			Recall		
		Internal Test w/ 95% CI	External Test #1 w/ 95% CI	External 2 Test #2 w/ 95% CI	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)													
	Male	0.779 (0.779 – 0.780)	0.724 (0.722 – 0.726)	0.676 (0.676 – 0.677)	0.797	0.742	0.507	0.419	0.363	0.359	0.534	0.570	0.859
	Female	0.768 (0.766 – 0.769)	0.763 (0.761 – 0.764)	0.666 (0.664 – 0.666)	0.712	0.788	0.570	0.333	0.384	0.312	0.676	0.537	0.711
Using Sensitive Attributes (M2)													
	Male	0.778 (0.777 – 0.778)	0.731 (0.730 – 0.734)	0.681 (0.679 – 0.682)	0.790	0.751	0.618	0.409	0.370	0.410	0.563	0.535	0.678
	Female	0.765 (0.764 – 0.766)	0.764 (0.764 – 0.765)	0.674 (0.671 – 0.676)	0.759	0.658	0.617	0.370	0.294	0.334	0.581	0.797	0.657
Separate Models for Each Group (M3)													
	Male	0.77 (0.768 – 0.771)	0.713 (0.711 – 0.716)	0.675 (0.673 – 0.675)	0.778	0.634	0.564	0.388	0.289	0.38	0.552	0.702	0.761
	Female	0.765 (0.764 – 0.766)	0.755 (0.752 – 0.756)	0.67 (0.667 – 0.671)	0.672	0.806	0.675	0.31	0.41	0.362	0.743	0.459	0.53
Race													
Not Using Sensitive Attributes (M1)													
	White	0.776 (0.775 – 0.776)	0.750 (0.749-0.750)	0.675 (0.674 – 0.675)	0.813	0.695	0.611	0.420	0.301	0.349	0.494	0.718	0.676
	NonWhite	0.762 (0.762-0.764)	0.747 (0.746-0.748)	0.666 (0.666 – 0.667)	0.711	0.762	0.629	0.364	0.371	0.388	0.667	0.556	0.614
Using Sensitive Attributes (M2)													
	White	0.774 (0.773-0.774)	0.751 (0.751-0.752)	0.676 (0.676 – 0.678)	0.788	0.703	0.637	0.384	0.312	0.365	0.564	0.744	0.648
	NonWhite	0.761 (0.760-0.761)	0.750 (0.749 – 0.751)	0.676 (0.674-0.677)	0.711	0.762	0.639	0.364	0.366	0.398	0.671	0.529	0.614
Separate Models for Each Group (M3)													
	White	0.766 (0.765-0.766)	0.755 (0.754-0.756)	0.665 (0.664-0.666)	0.761	0.799	0.584	0.352	0.362	0.336	0.612	0.538	0.711

	NonWhite	0.738 (0.387-0.739)	0.733 (0.731-0.733)	0.681 (0.680-0.682)	0.679	0.607	0.577	0.333	0.274	0.368	0.662	0.785	0.758
Ethnicity													
Not Using Sensitive Attributes (M1)													
	NonHispanic	0.770 (0.770-0.772)	0.747 (0.746-0.748)	0.666 (0.665-0.666)	0.785	0.766	0.632	0.396	0.373	0.357	0.519	0.544	0.617
	Hispanic	0.821 (0.819-0.822)	0.808 (0.802- 0.809)	0.689 (0.686-0.690)	0.733	0.722	0.539	0.407	0.231	0.436	0.857	1.000	0.875
Using Sensitive Attributes (M2)													
	NonHispanic	0.769 (0.768-0.769)	0.751 (0.750-0.752)	0.667 (0.667-0.668)	0.772	0.763	0.631	0.382	0.363	0.356	0.570	0.514	0.617
	Hispanic	0.814 (0.811 – 0.815)	0.798 (0.797-0.799)	0.700 (0.698-0.701)	0.719	0.750	0.542	0.390	0.250	0.441	0.821	1.000	0.938
Separate Models for Each Group (M3)													
	NonHispanic	0.767 (0.766-0.767)	0.754 (0.752-0.754)	0.657 (0.656-0.658)	0.776	0.762	0.58	0.384	0.37	0.332	0.538	0.559	0.693
	Hispanic	0.722 (0.720 – 0.723)	0.556 (0.555-0.558)	0.677 (0.676-0.678)	0.664	0.889	0.565	0.328	0.333	0.466	0.714	0.333	0.777