Ensembles

Main ensembles

Main ensembles considered are: unweighted mean/median; mean/median weighted by inverse scaled relative WIS.

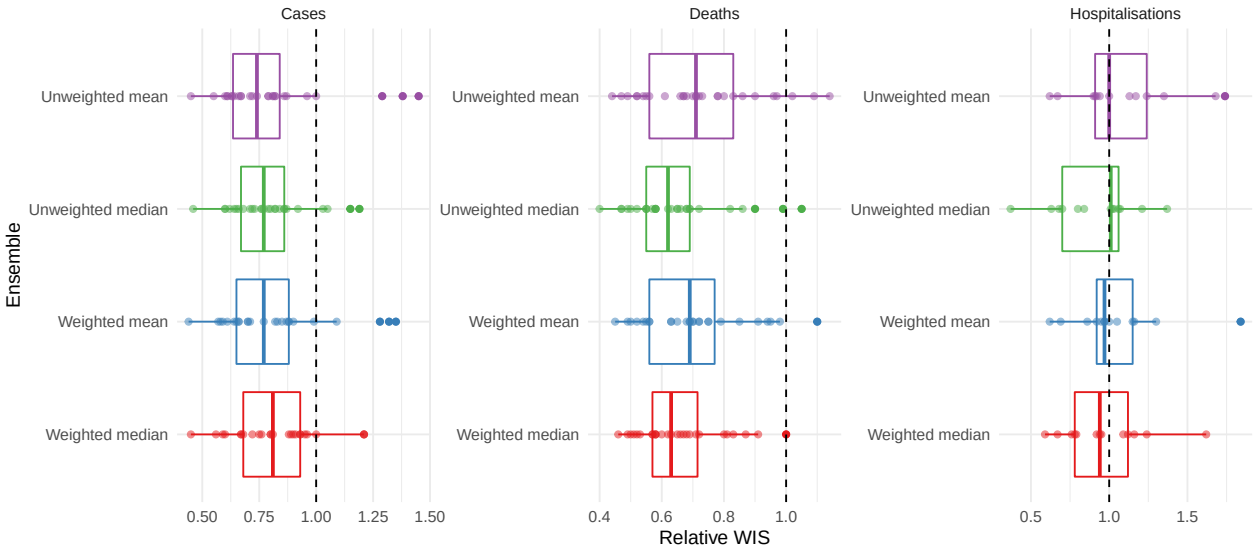


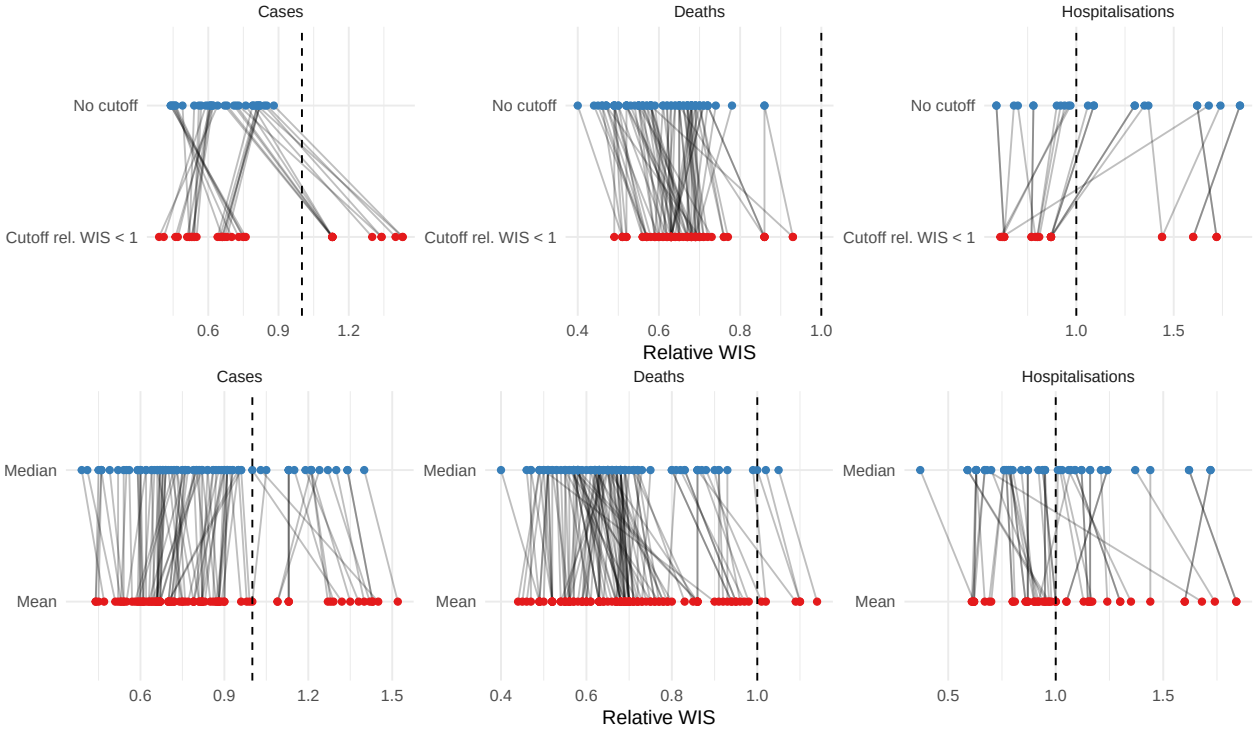
Figure 1: Predictive performance at the 2 week horizon. Main ensembles, using all component models and considering all of the history. Each point is one country.

Variations on main ensembles

We further considered: a *cutoff* (i.e., creating only ensembles from models that had relative scaled WIS < 1 across all forecast horizons for a given target/country); and *limited history* (i.e., only considering the last 10 weeks of scores for the weighted ensemble, as opposed to the whole history). Each point in the countries below represent one country/method/target triplet at the 2-week horizon, with only one aspect of the ensemble varying between the top and bottom rows, connected by lines.

Table 1: Predictive performance of main ensembles, as measured by the scaled relative WIS.

Horizon	Weighted mean	Weighted median	Unweighted mean	Unweighted median
Cases				
1 week	0.60	0.64	0.61	0.64
2 weeks	0.73	0.74	0.74	0.74
3 weeks	0.89	0.79	0.90	0.79
4 weeks	1.21	0.86	1.22	0.85
Deaths				
1 week	0.62	0.59	0.63	0.59
2 weeks	0.56	0.54	0.56	0.52
3 weeks	0.62	0.54	0.62	0.52
4 weeks	0.74	0.60	0.74	0.58
Hospitalisations				
1 week	0.93	0.83	1.13	0.83
2 weeks	0.99	0.87	1.25	0.86
3 weeks	1.08	0.94	1.35	0.90
4 weeks	1.09	0.99	1.31	0.94



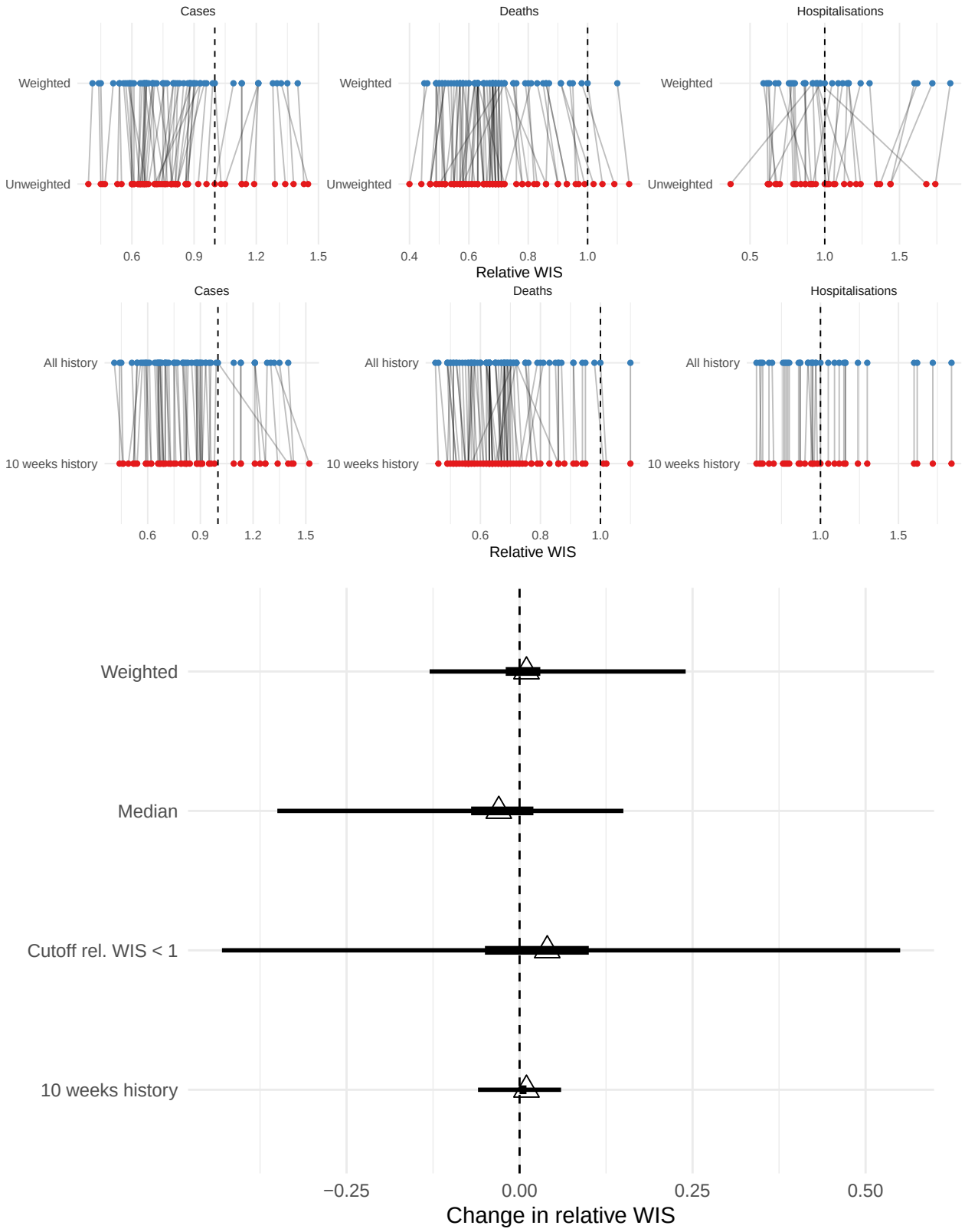


Figure 2: Aggregate differences between ensemble variants; shown are the central 48% and 96% intervals (thick/thin lines) and means of the distribution of difference in rel. WIS across all methods/countries.