

Supplementary material for: When do we need multiple infectious  
disease models? Agreement between projection rank and magnitude  
in a multi-model setting

La Keisha Wade-Malone, The Pennsylvania State University

Emily Howerton, The Pennsylvania State University

William J.M. Probert, University of Oxford

Michael C. Runge, U.S. Geological Survey

Cecile Viboud, National Institutes of Health

Katriona Shea, The Pennsylvania State University

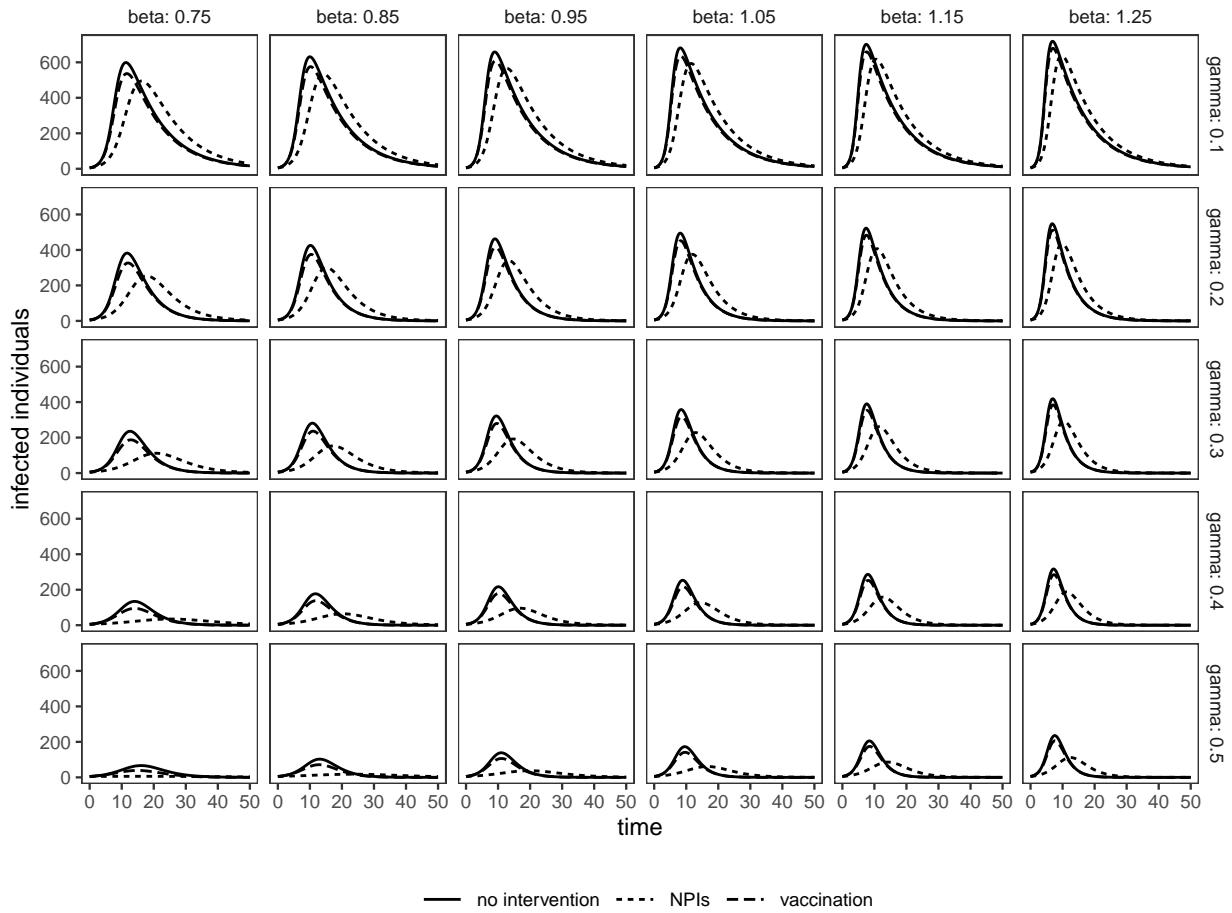


Fig. S1: Susceptible-Infected-Recovered (SIR) dynamics of infected individuals for varying transmission rate ( $\beta$ ) and recovery rate ( $\gamma$ ) parameters. Dynamics are shown for three scenarios: no intervention (solid line), non-pharmaceutical interventions (NPIs) that reduce transmission by 30% (short dashed line), and vaccination of approximately 1% of the susceptible population per day (long dashed line). Data from SIR model simulations.

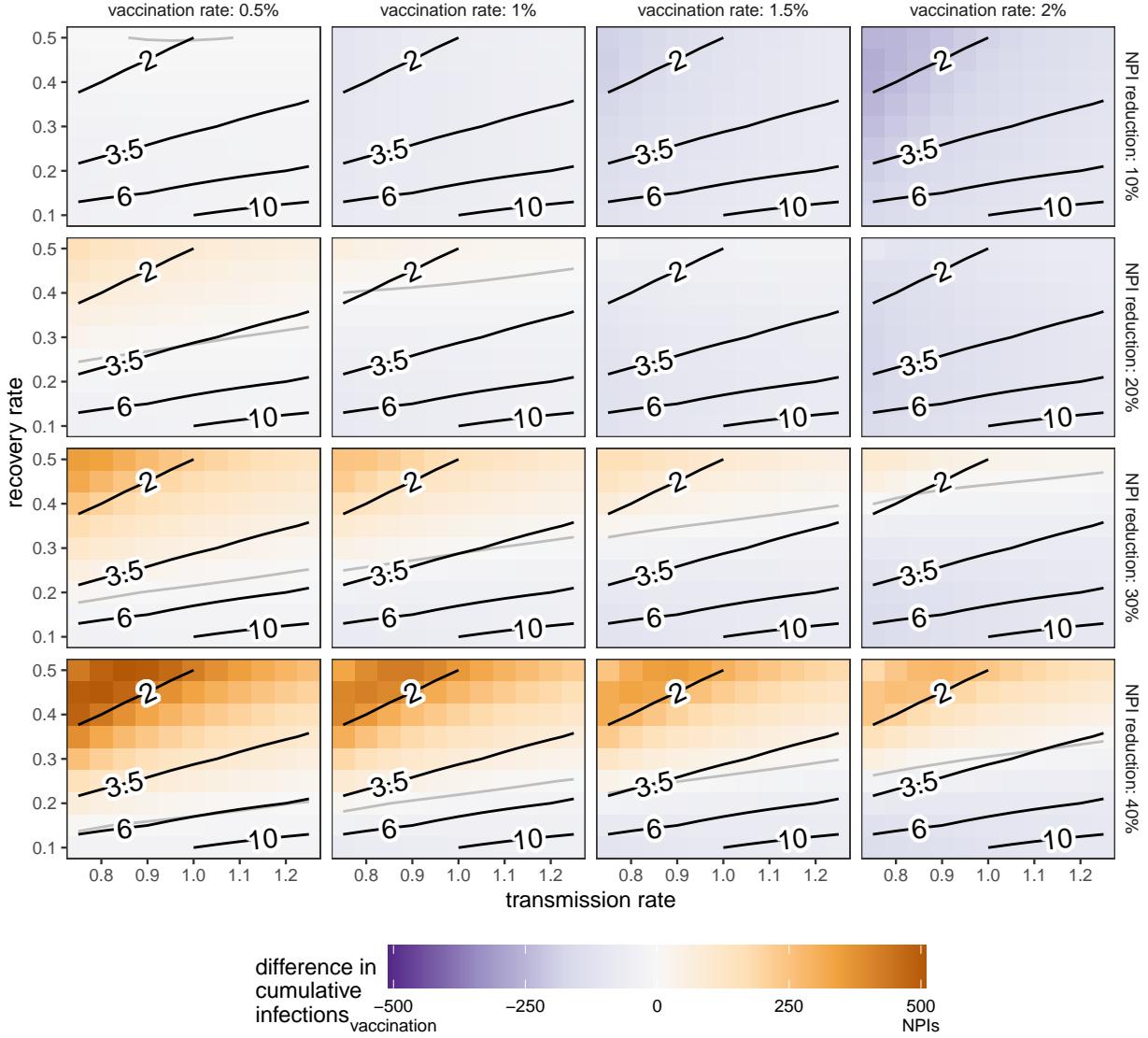


Fig. S2: Results from Fig. 1B repeated for varying intervention efficacies (vaccination rate ranging from 0.5% to 2% of the population per day, and NPI reduction of the transmission rate ranging from 10% to 40%). Each panel shows a large part of the space of possible models across biological uncertainties (transmission rate,  $\beta$ , and recovery rate,  $\gamma$ ). Each position on the graph represents an individual model, the color of the tile represents the recommended intervention (orange: non-pharmaceutical interventions, NPIs; purple: vaccination), and the intensity of the color is the magnitude of the projected difference in cumulative infections. Colors have been rescaled from Fig. 1. The gray line shows where the difference in cumulative infections is 0, or where neither intervention is recommended over the other. Black contours show sample values of individual model  $R_0 = 2, 3.5, 6$ , and 10. Data from SIR model simulations.

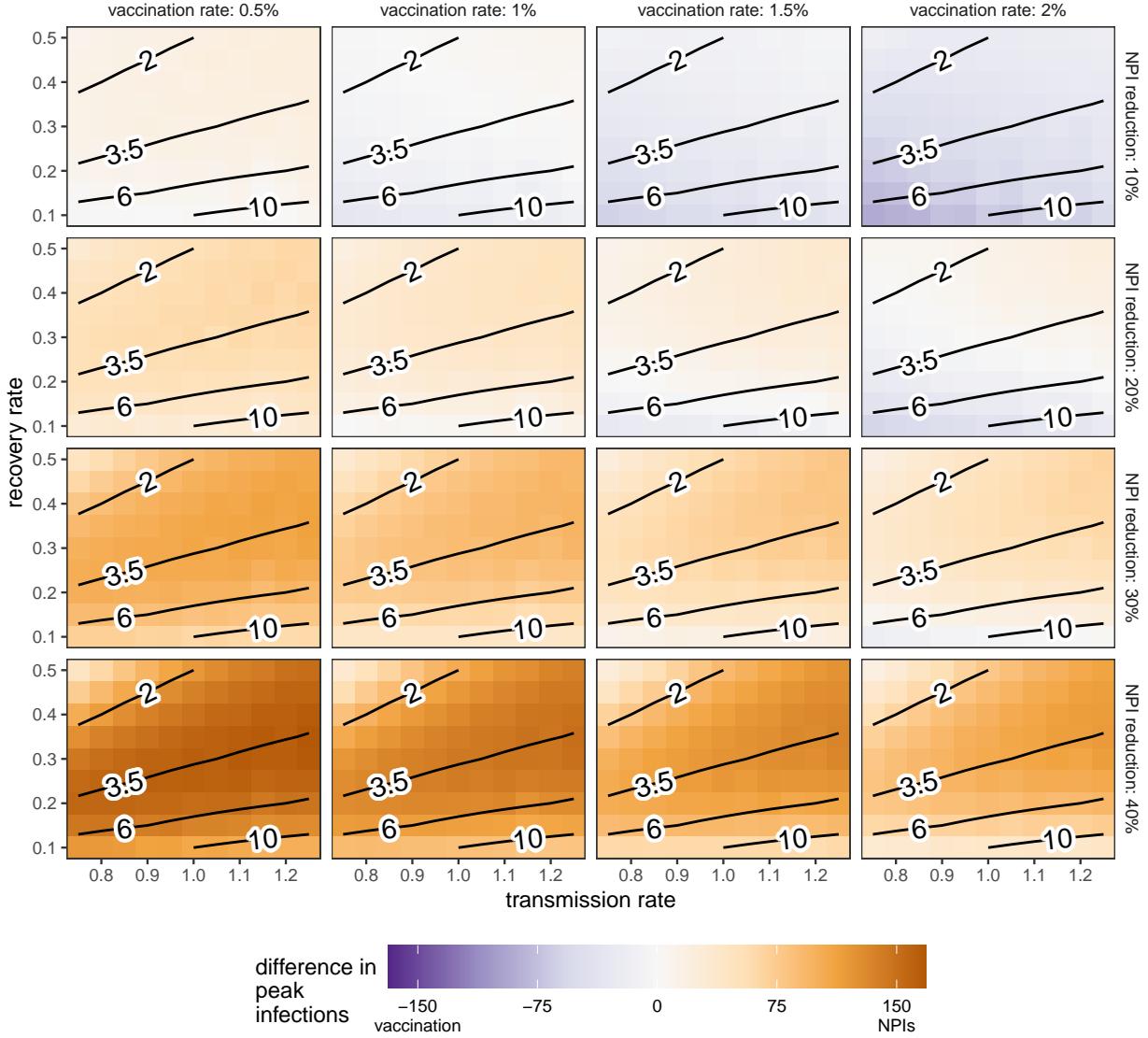


Fig. S3: Model results for minimizing peak infections. Results are shown for varying intervention efficacies (vaccination rate ranging from 0.5% to 2% of the population per day, and NPI reduction of the transmission rate ranging from 10% to 40%). Each panel shows a large part of the space of possible models across biological uncertainties (transmission rate,  $\beta$ , and recovery rate,  $\gamma$ ). Each position on the graph represents an individual model, the color of the tile represents the recommended intervention (orange: non-pharmaceutical interventions, NPIs; purple: vaccination), and the intensity of the color is the magnitude of the projected difference in peak infections. Colors have been rescaled from Figure 1. The gray line shows where the difference in peak infections is 0, or where neither intervention is recommended over the other. Black contours show sample values of individual model  $R_0 = 2, 3.5, 6$ , and  $10$ . Data from SIR model simulations.

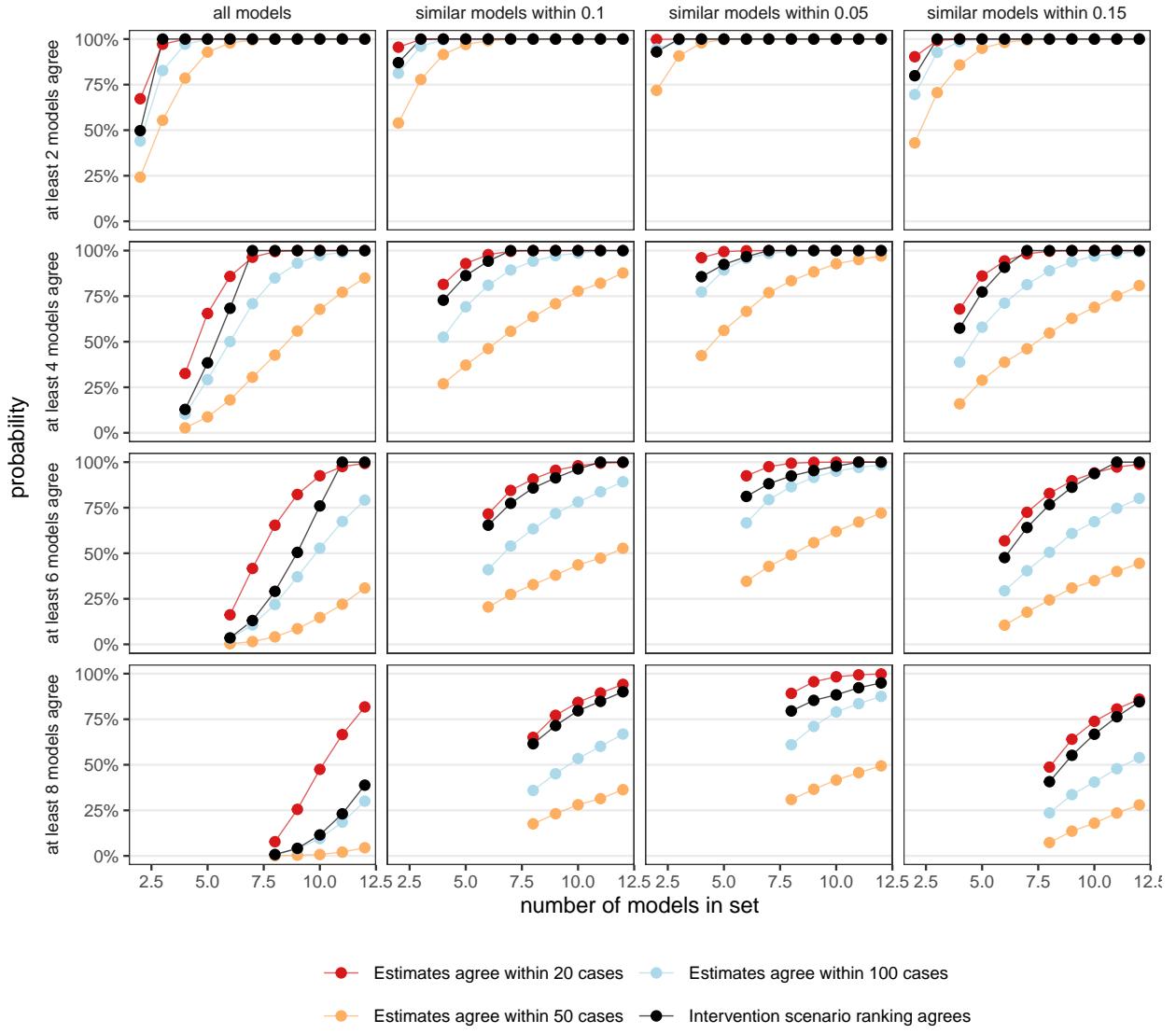


Fig. S4: Probability of models agreeing about intervention scenario ranking or estimates of cumulative infections without intervention for different sampling schemes (columns) and number of models agreeing (rows). Probabilities are calculated for a set of models of size two to size twelve, either when choosing randomly among all models or among models with “similar” assumptions, defined as those with transmission and recovery rate assumptions within  $\pm 0.1$ ,  $\pm 0.05$  or  $\pm 0.15$  of a given model. Agreement probability of at least 2, 4, 6, or 8 models is shown for estimates of cumulative cases within 20 infections (2% of population, red), within 50 infections (5% of population, yellow), within 100 infections (10% of population, blue), or intervention scenario ranking (black). Data from SIR model simulations.

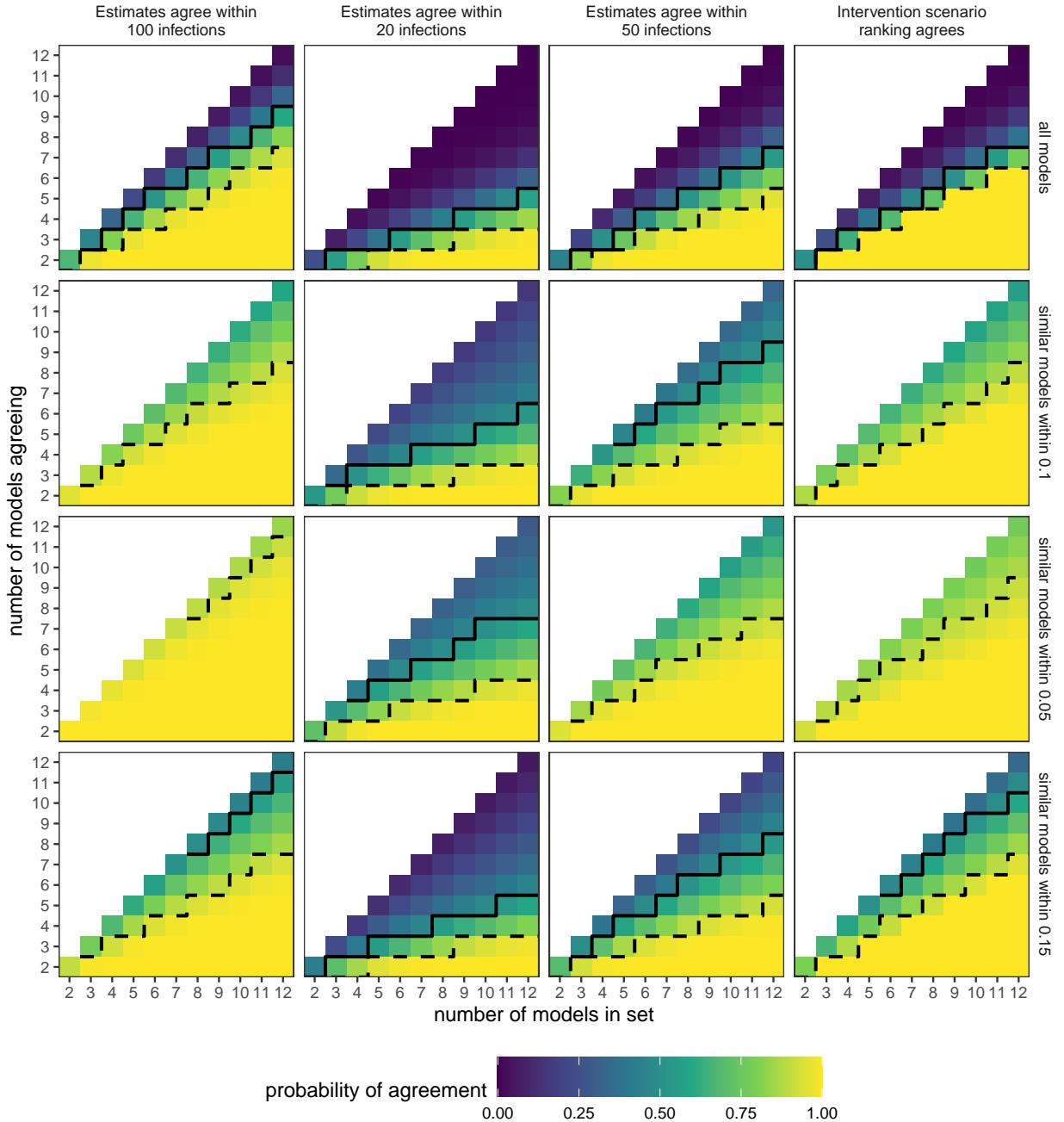


Fig. S5: Probability of models agreeing about intervention scenario ranking or estimates of cumulative infections without intervention (columns) for different degrees of model similarity (rows). Probabilities are calculated for a set of models of size two to size twelve, either when choosing randomly among all models or among models with “similar” assumptions, defined as those with transmission and recovery rate assumptions within  $\pm 0.1$ ,  $\pm 0.05$  or  $\pm 0.15$  of a given model. Agreement probability of at least 2, 4, 6, or 8 models is shown for estimates of cumulative cases within 20 infections (2% of population), within 50 infections (5% of population), within 100 infections (10% of population), or intervention scenario ranking. Breakpoints where agreement probability is at least 0.5 is shown with a solid black line and agreement probability is at least 0.9 is shown with a dashed black line. Data from SIR model simulations.

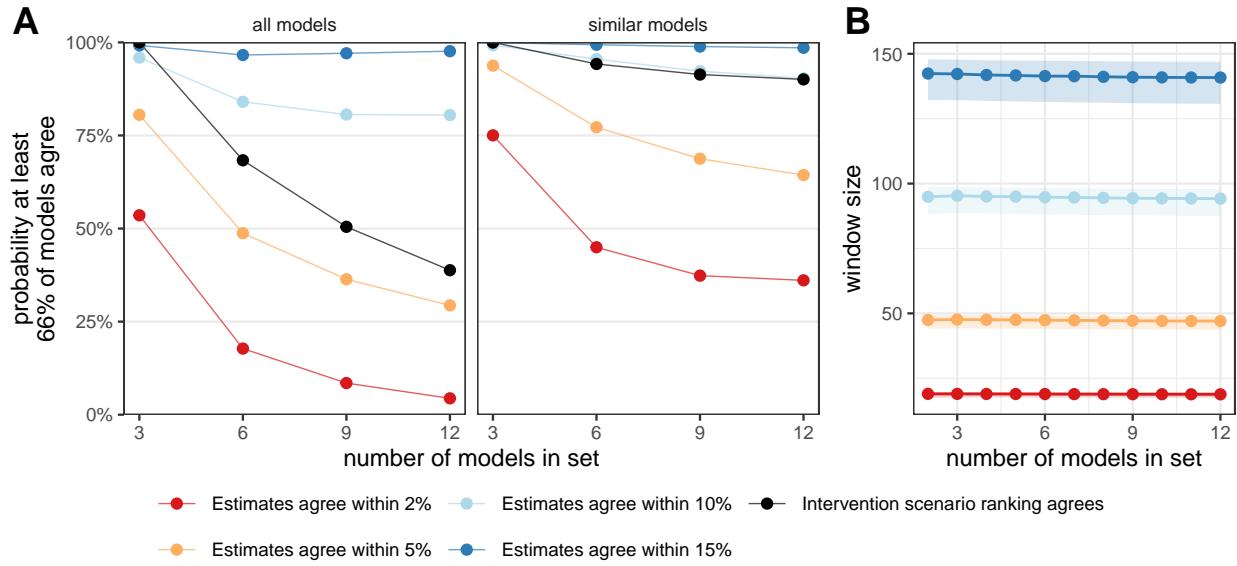


Fig. S6: (A) Probability of at least 66% of models agreeing for a relative definition of agreement. Here, agreement probability for projection magnitude is calculated relative to the projection size, i.e., agreement constitutes the maximum number of estimates from a set with cumulative cases falling within a window of 2% of the projected magnitude (red), within 5% of the projected magnitude (yellow), within 10% of the projected magnitude (light blue), or within 15% of the projected magnitude (dark blue). Agreement about recommended intervention is shown for reference (black). Probabilities are calculated for agreement of 2 models out of a set of 3, 4 out of a set of 6, 6 out of set of 8, and 8 out of a set of 12. All probabilities were calculated both when choosing randomly among all models or among models with “similar” assumptions. Similar models are defined as those with transmission and recovery rate assumptions within  $\pm 0.1$  of a given model. (B) The size of the window within which cumulative case estimates are considered to agree, for 2% of the projected magnitude (red), 5% (yellow), 10% (light blue), or 15% (dark blue). The points show the average across all randomly drawn sets of models, and the ribbon show the 25th and 75th quantiles of this distribution. Data from SIR model simulations.

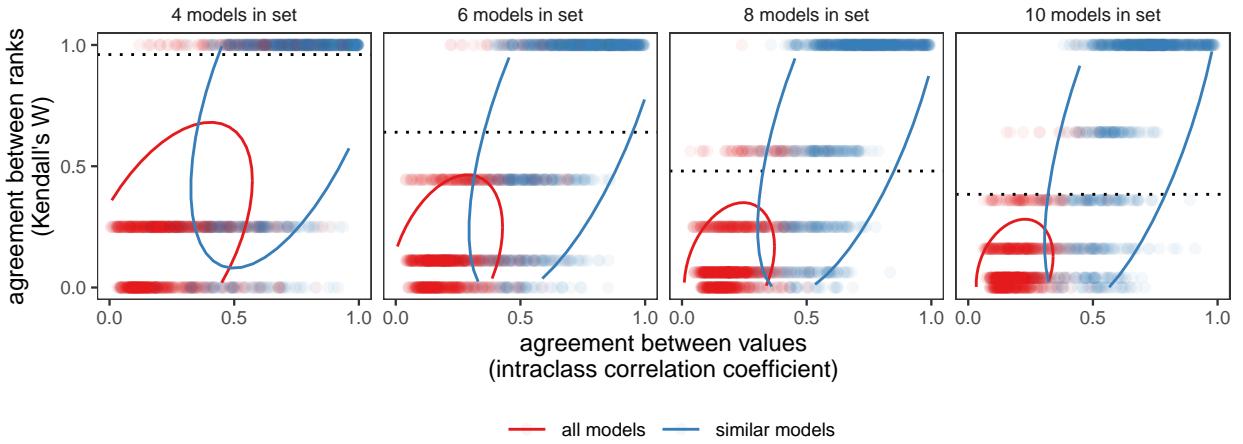


Fig. S7: Results from Fig. 3 replicated with agreement assessed for only two intervention scenarios. For a varying number of models in a set, agreement is measured for projection magnitude across intervention scenarios (using intraclass correlation coefficient, ICC) and ranking of intervention scenarios (using Kendall's W). Results are shown when models in the set come from all possible models considered (red), and when models are similar (blue). Each point represents one set of randomly selected models from either all possible models considered (red) or a neighborhood of “similar” models (blue, those with transmission and recovery rate assumptions within  $\pm 0.1$  of a given model). Ellipses show the area within which 95% of points fall. Dotted horizontal lines show agreement that is significant with a p-value of 0.05. Data from SIR model simulations.

### Round 1 (horizon: 26 weeks)

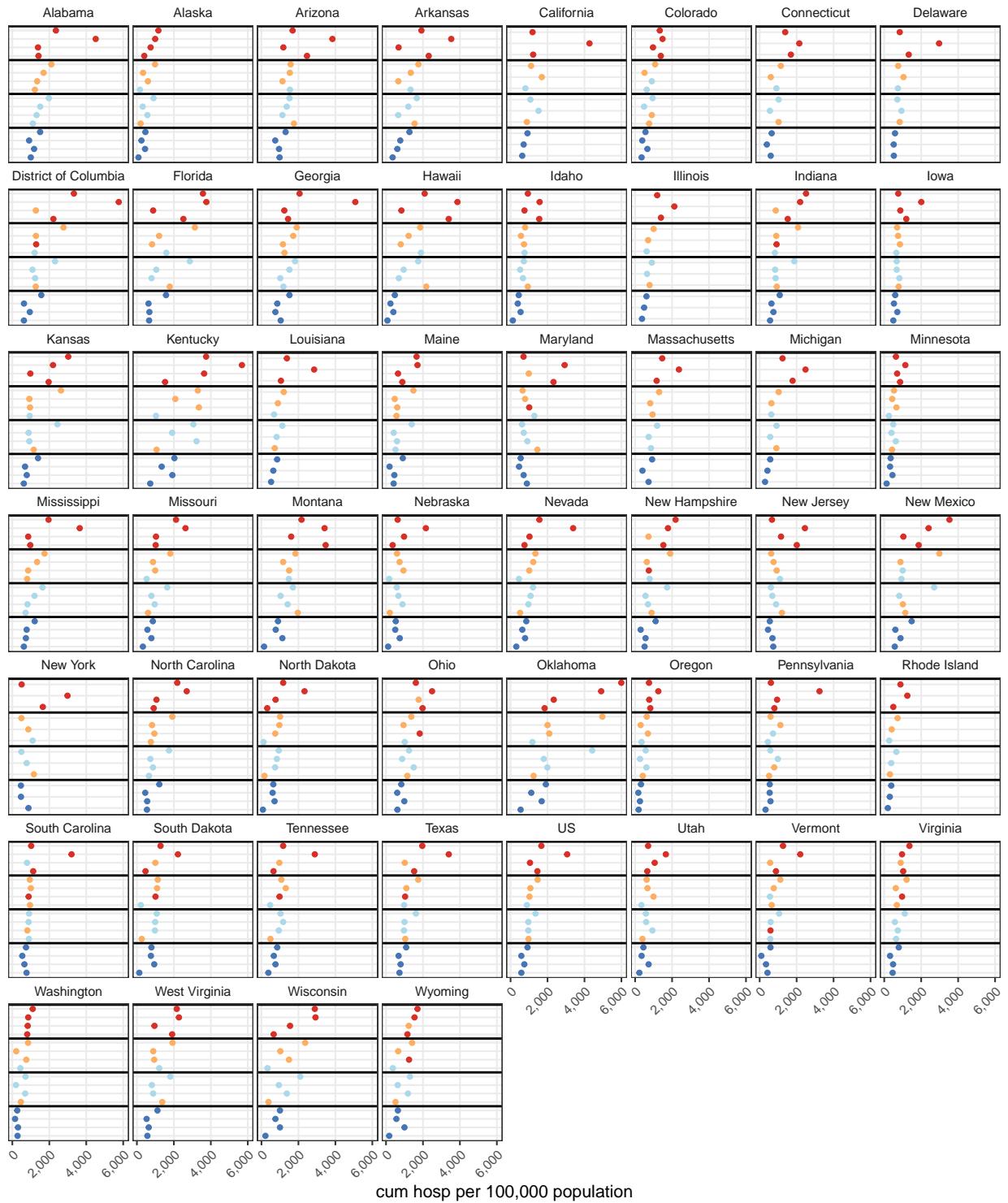


Fig. S8: COVID-19 Scenario Modeling Hub projections of cumulative hospitalizations for Round 1. Projections are collected from 4 models for four scenarios. Across locations (panels), projections are shown for each scenario (separated by horizontal black lines) from each model (individual points), points are colored based on the rank of that scenario for each model (blue is lowest projection, and red is highest projection). Data from [https://github.com/midas-network/covid19-scenario-hub\\_evaluation](https://github.com/midas-network/covid19-scenario-hub_evaluation), <https://zenodo.org/record/8415147>.

### Round 2 (horizon: 26 weeks)



Fig. S9: COVID-19 Scenario Modeling Hub projections of cumulative hospitalizations for Round 2. Projections are collected from 5 models for four scenarios. Across locations (panels), projections are shown for each scenario (separated by horizontal black lines) from each model (individual points), points are colored based on the rank of that scenario for each model (blue is lowest projection, and red is highest projection). Data from [https://github.com/midas-network/covid19-scenario-hub\\_evaluation](https://github.com/midas-network/covid19-scenario-hub_evaluation), <https://zenodo.org/record/8415147>.

### Round 3 (horizon: 26 weeks)

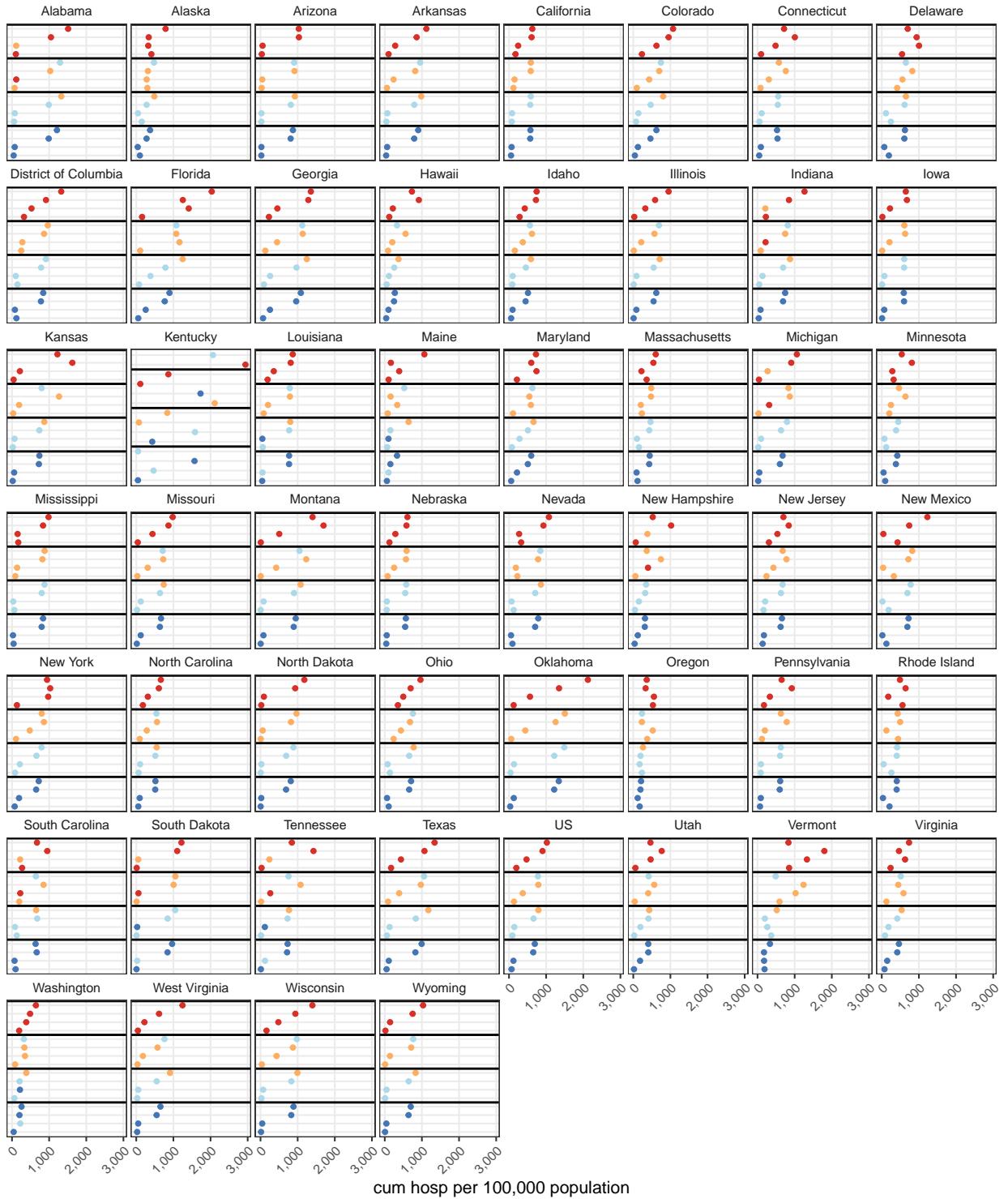


Fig. S10: COVID-19 Scenario Modeling Hub projections of cumulative hospitalizations for Round 3. Projections are collected from 4 models for four scenarios. Across locations (panels), projections are shown for each scenario (separated by horizontal black lines) from each model (individual points), points are colored based on the rank of that scenario for each model (blue is lowest projection, and red is highest projection). Data from [https://github.com/midas-network/covid19-scenario-hub\\_evaluation](https://github.com/midas-network/covid19-scenario-hub_evaluation), <https://zenodo.org/record/8415147>.

Round 4 (horizon: 26 weeks)

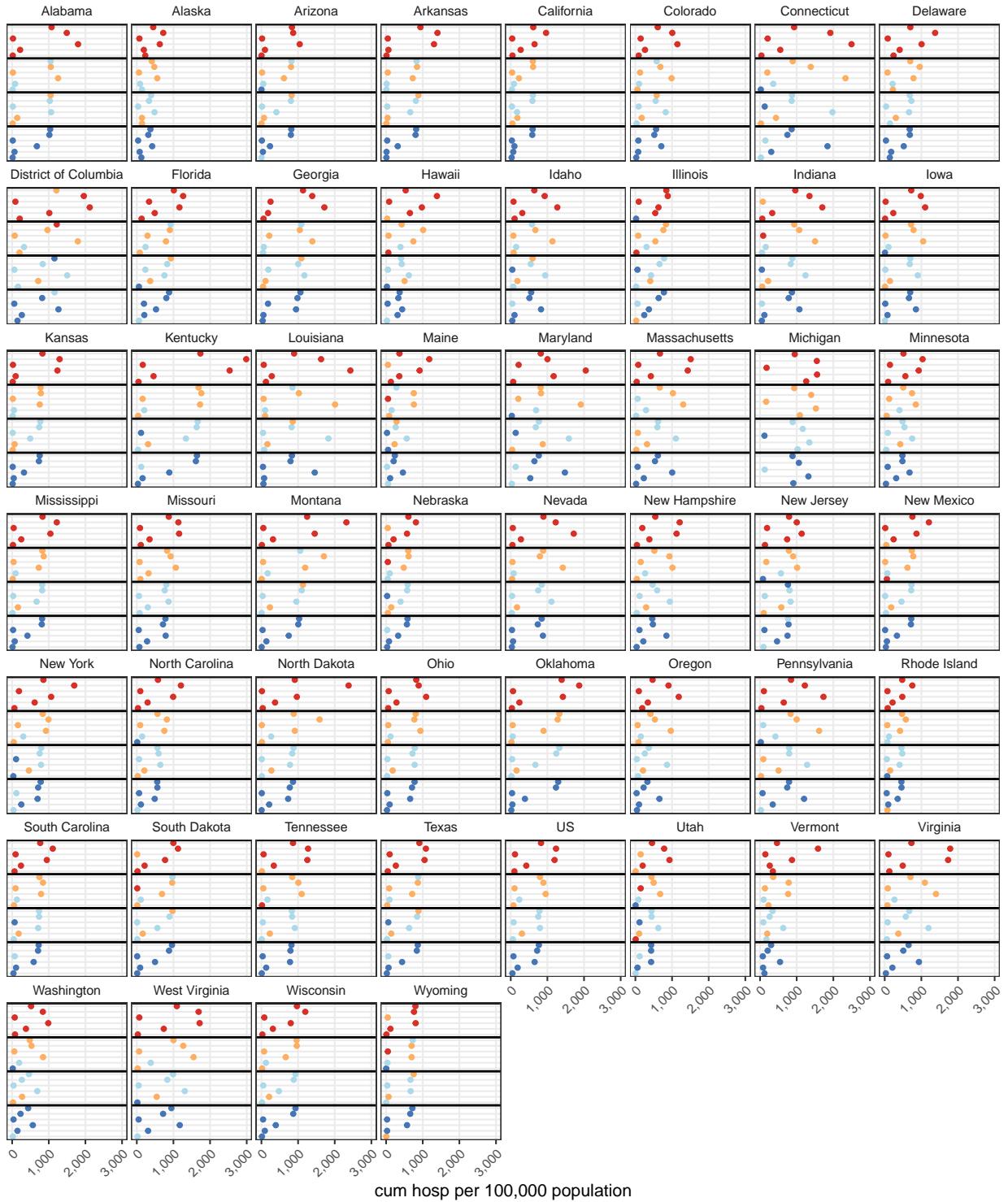


Fig. S11: COVID-19 Scenario Modeling Hub projections of cumulative hospitalizations for Round 4. Projections are collected from 6 models for four scenarios. Across locations (panels), projections are shown for each scenario (separated by horizontal black lines) from each model (individual points), points are colored based on the rank of that scenario for each model (blue is lowest projection, and red is highest projection). Data from [https://github.com/midas-network/covid19-scenario-hub\\_evaluation](https://github.com/midas-network/covid19-scenario-hub_evaluation), <https://zenodo.org/record/8415147>.

Round 5 (horizon: 26 weeks)

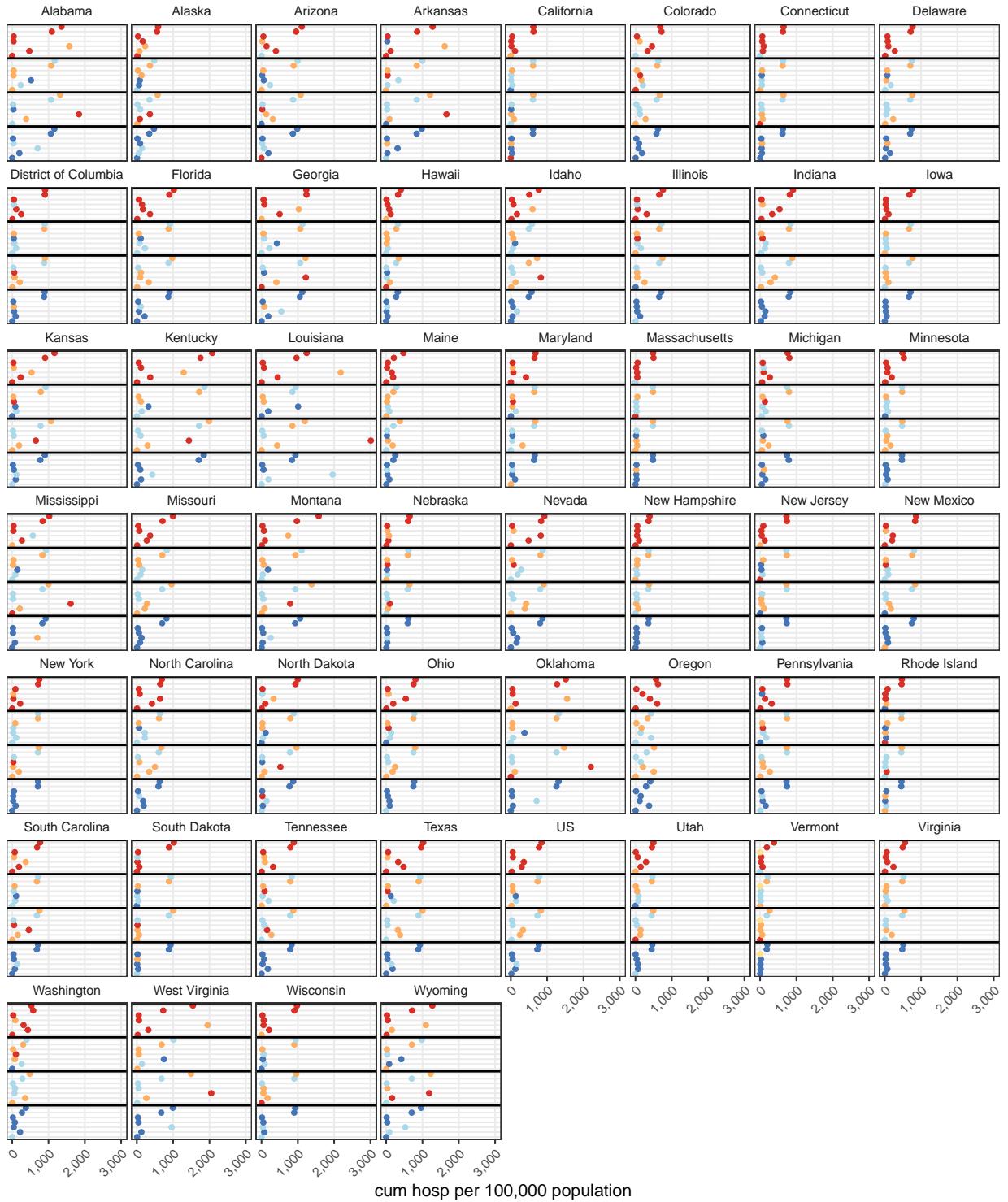


Fig. S12: COVID-19 Scenario Modeling Hub projections of cumulative hospitalizations for Round 5. Projections are collected from 7 models for four scenarios. Across locations (panels), projections are shown for each scenario (separated by horizontal black lines) from each model (individual points), points are colored based on the rank of that scenario for each model (blue is lowest projection, and red is highest projection). Data from [https://github.com/midas-network/covid19-scenario-hub\\_evaluation](https://github.com/midas-network/covid19-scenario-hub_evaluation), <https://zenodo.org/record/8415147>.

Round 6 (horizon: 26 weeks)

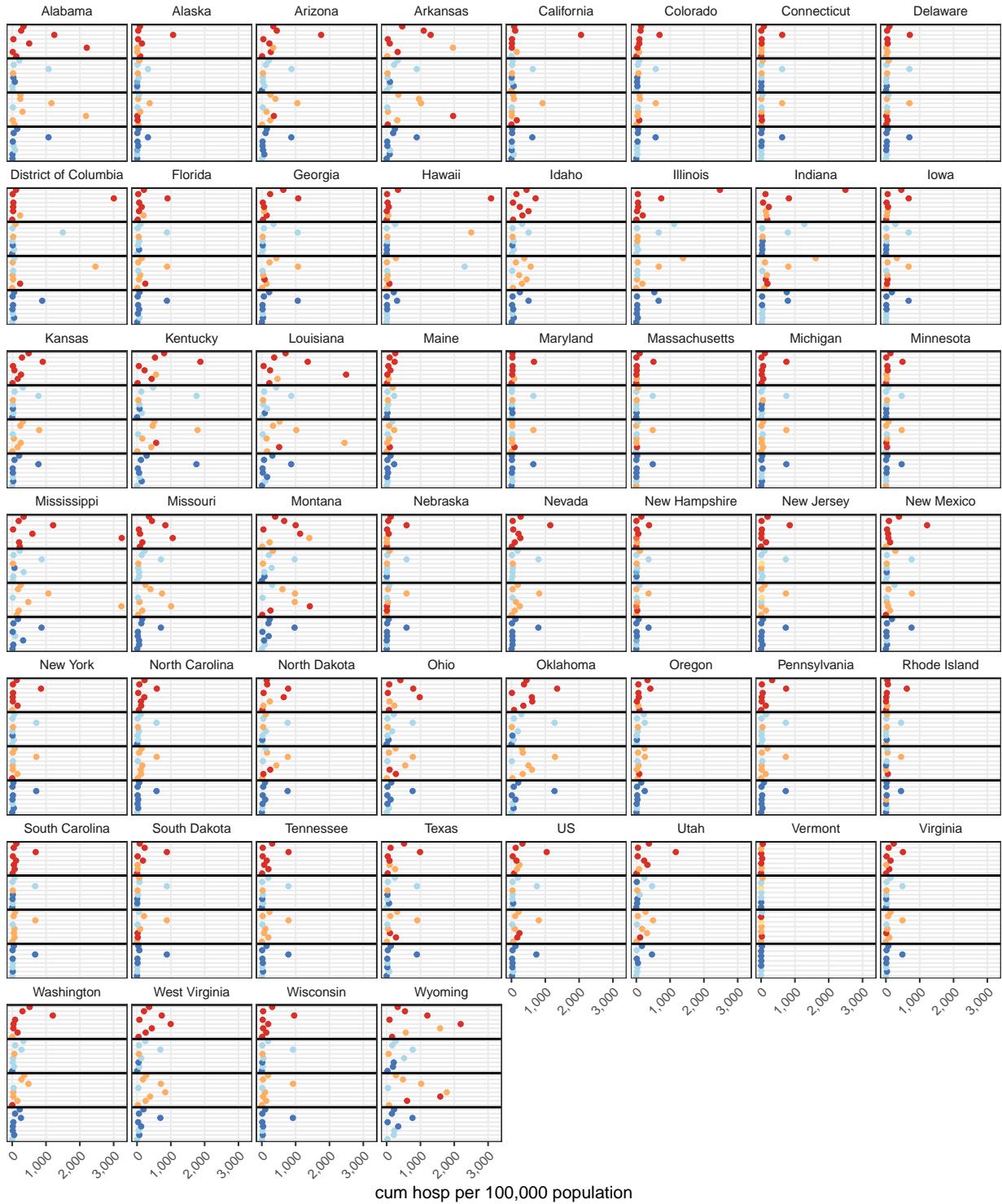


Fig. S13: COVID-19 Scenario Modeling Hub projections of cumulative hospitalizations for Round 6. Projections are collected from 8 models for four scenarios. Across locations (panels), projections are shown for each scenario (separated by horizontal black lines) from each model (individual points), points are colored based on the rank of that scenario for each model (blue is lowest projection, and red is highest projection). Data from [https://github.com/midas-network/covid19-scenario-hub\\_evaluation](https://github.com/midas-network/covid19-scenario-hub_evaluation), <https://zenodo.org/record/8415147>.

Round 7 (horizon: 26 weeks)



Fig. S14: COVID-19 Scenario Modeling Hub projections of cumulative hospitalizations for Round 7. Projections are collected from 8 models for four scenarios. Across locations (panels), projections are shown for each scenario (separated by horizontal black lines) from each model (individual points), points are colored based on the rank of that scenario for each model (blue is lowest projection, and red is highest projection). Data from [https://github.com/midas-network/covid19-scenario-hub\\_evaluation](https://github.com/midas-network/covid19-scenario-hub_evaluation), <https://zenodo.org/record/8415147>.

### Round 9 (horizon: 26 weeks)

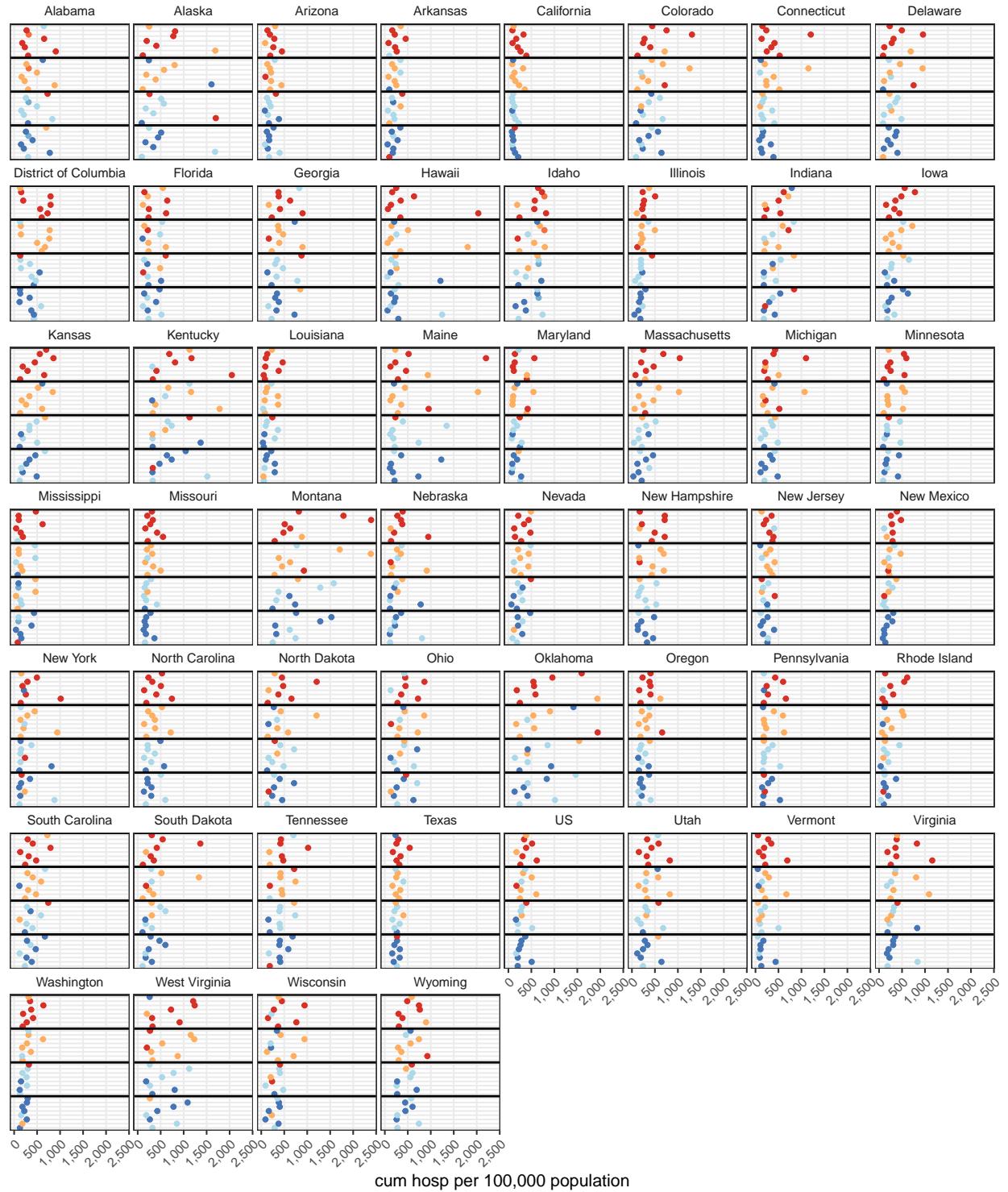


Fig. S15: COVID-19 Scenario Modeling Hub projections of cumulative hospitalizations for Round 9. Projections are collected from 8 models for four scenarios. Across locations (panels), projections are shown for each scenario (separated by horizontal black lines) from each model (individual points), points are colored based on the rank of that scenario for each model (blue is lowest projection, and red is highest projection). Data from [https://github.com/midas-network/covid19-scenario-hub\\_evaluation](https://github.com/midas-network/covid19-scenario-hub_evaluation), <https://zenodo.org/record/8415147>.

### Round 11 (horizon: 12 weeks)



Fig. S16: COVID-19 Scenario Modeling Hub projections of cumulative hospitalizations for Round 11. Projections are collected from 6 models for four scenarios. Across locations (panels), projections are shown for each scenario (separated by horizontal black lines) from each model (individual points), points are colored based on the rank of that scenario for each model (blue is lowest projection, and red is highest projection). Data from [https://github.com/midas-network/covid19-scenario-hub\\_evaluation](https://github.com/midas-network/covid19-scenario-hub_evaluation), <https://zenodo.org/record/8415147>.

### Round 12 (horizon: 12 weeks)

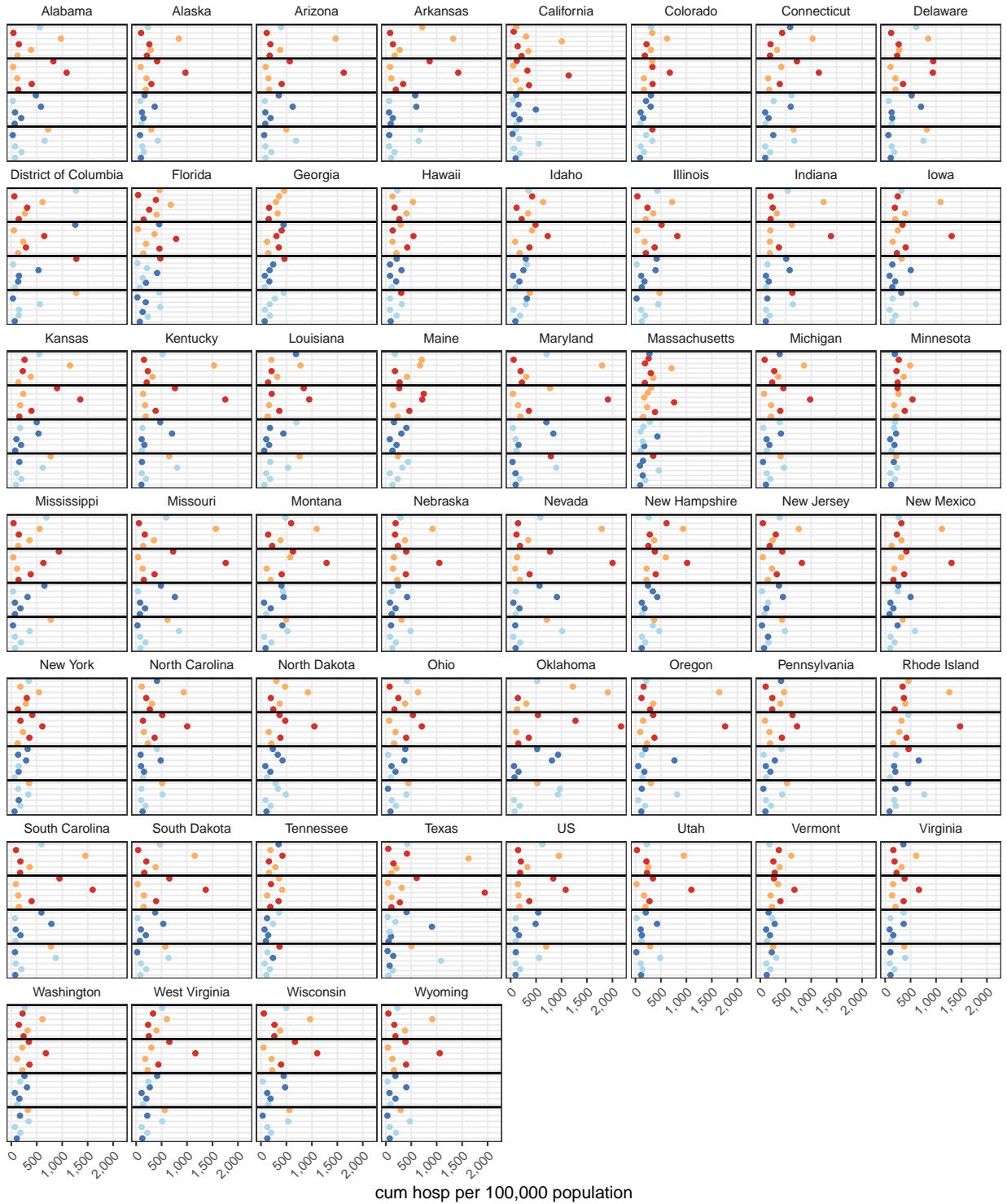


Fig. S17: COVID-19 Scenario Modeling Hub projections of cumulative hospitalizations for Round 12. Projections are collected from 6 models for four scenarios. Across locations (panels), projections are shown for each scenario (separated by horizontal black lines) from each model (individual points), points are colored based on the rank of that scenario for each model (blue is lowest projection, and red is highest projection). Data from [https://github.com/midas-network/covid19-scenario-hub\\_evaluation](https://github.com/midas-network/covid19-scenario-hub_evaluation), <https://zenodo.org/record/8415147>.

Round 13 (horizon: 52 weeks)

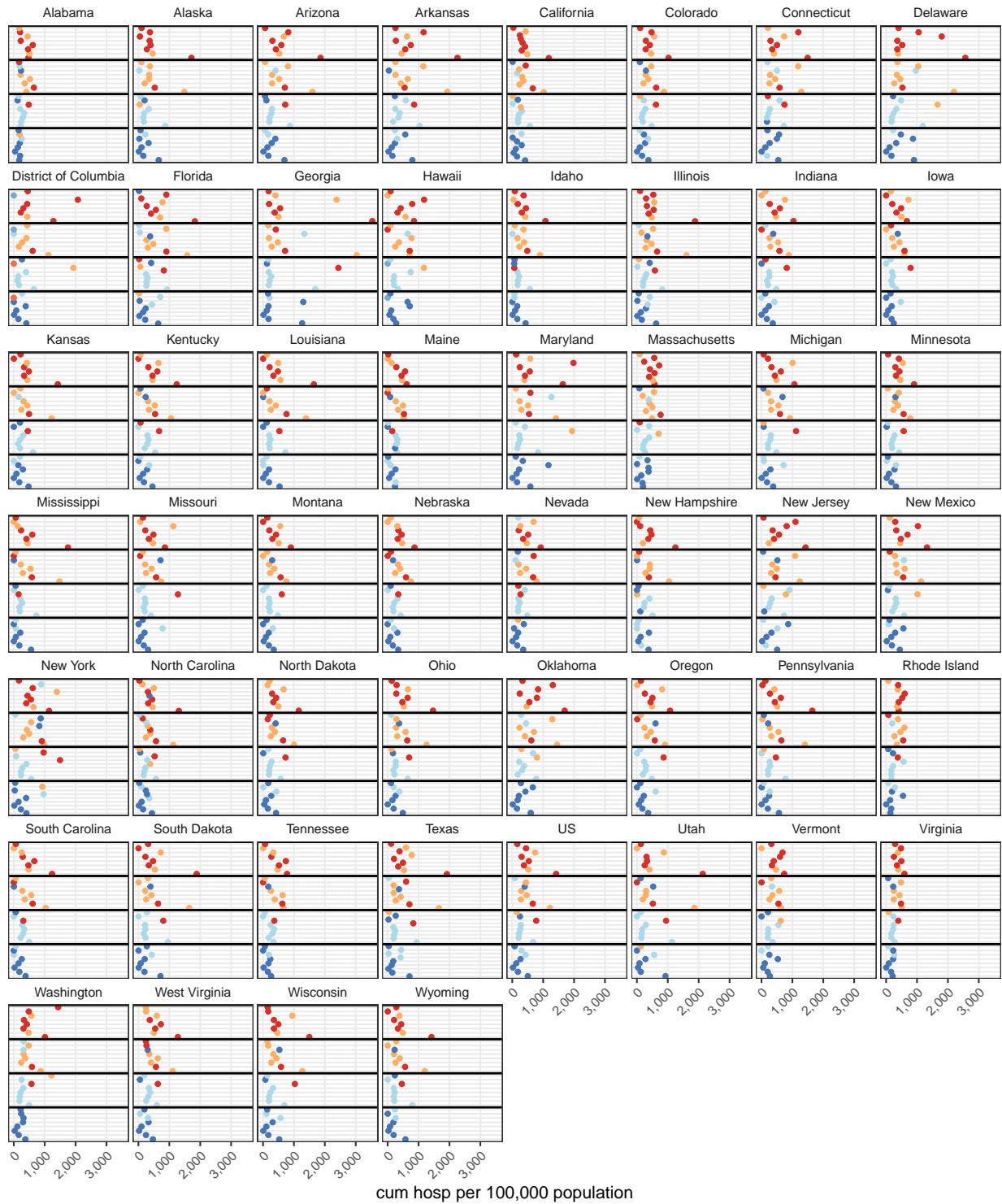


Fig. S18: COVID-19 Scenario Modeling Hub projections of cumulative hospitalizations for Round 13. Projections are collected from 8 models for four scenarios. Across locations (panels), projections are shown for each scenario (separated by horizontal black lines) from each model (individual points), points are colored based on the rank of that scenario for each model (blue is lowest projection, and red is highest projection). Data from [https://github.com/midas-network/covid19-scenario-hub\\_evaluation](https://github.com/midas-network/covid19-scenario-hub_evaluation), <https://zenodo.org/record/8415147>.

Round 14 (horizon: 50 weeks)

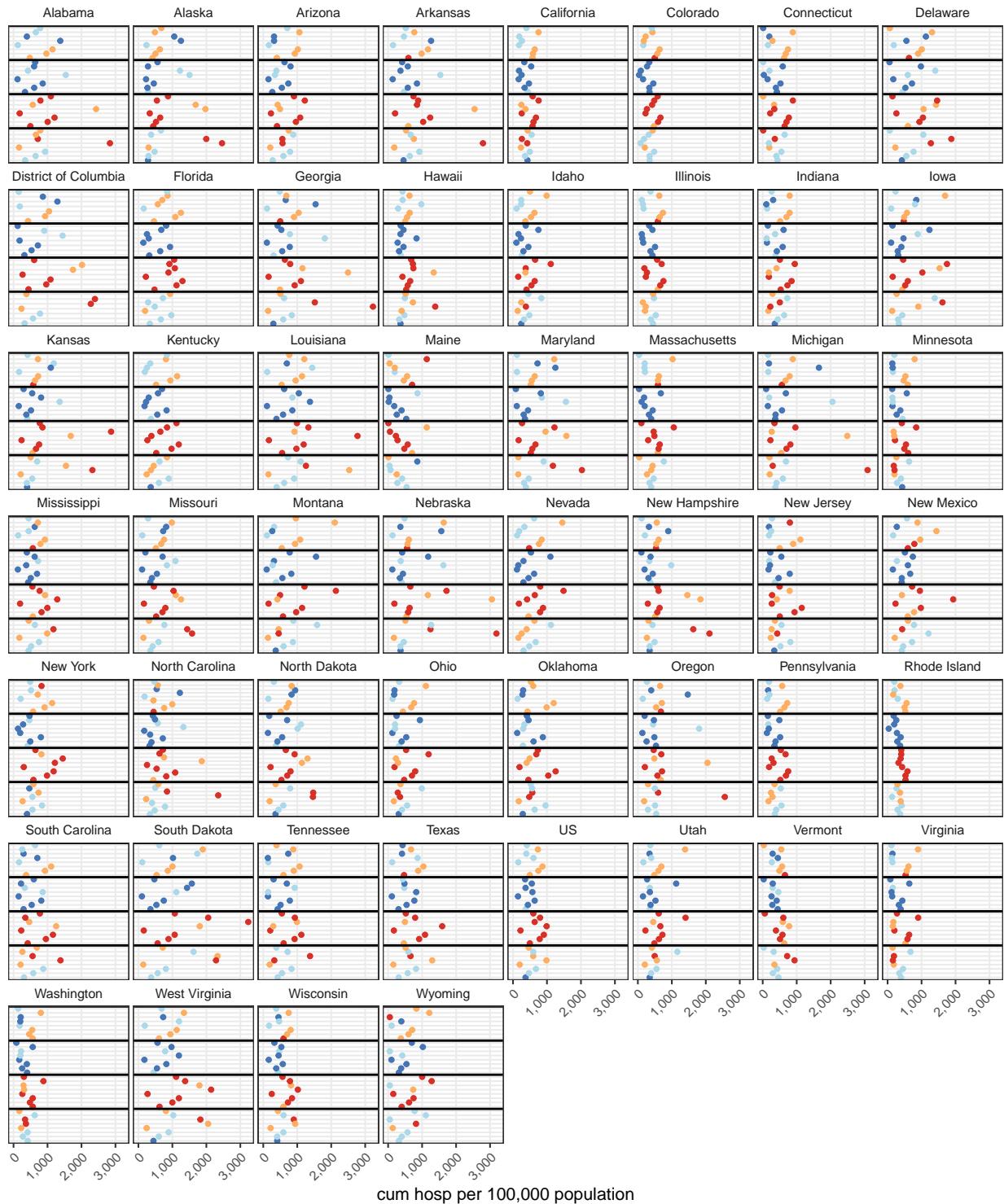


Fig. S19: COVID-19 Scenario Modeling Hub projections of cumulative hospitalizations for Round 14. Projections are collected from 8 models for four scenarios. Across locations (panels), projections are shown for each scenario (separated by horizontal black lines) from each model (individual points), points are colored based on the rank of that scenario for each model (blue is lowest projection, and red is highest projection). Data from [https://github.com/midas-network/covid19-scenario-hub\\_evaluation](https://github.com/midas-network/covid19-scenario-hub_evaluation), <https://zenodo.org/record/8415147>.

Round 15 (horizon: 40 weeks)

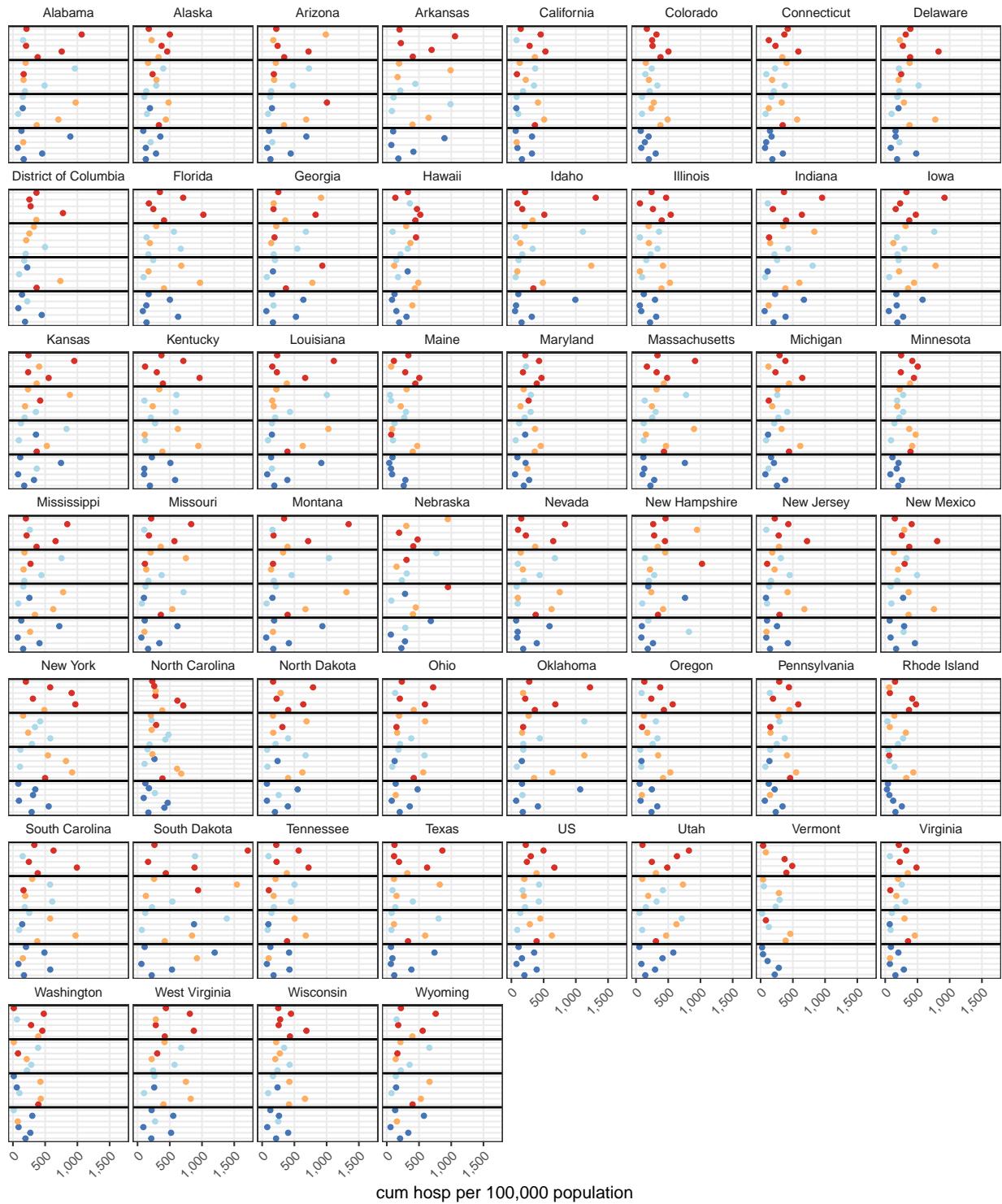


Fig. S20: COVID-19 Scenario Modeling Hub projections of cumulative hospitalizations for Round 15. Projections are collected from 6 models for four scenarios. Across locations (panels), projections are shown for each scenario (separated by horizontal black lines) from each model (individual points), points are colored based on the rank of that scenario for each model (blue is lowest projection, and red is highest projection). Data from [https://github.com/midas-network/covid19-scenario-hub\\_evaluation](https://github.com/midas-network/covid19-scenario-hub_evaluation), <https://zenodo.org/record/8415147>.

Round 16 (horizon: 26 weeks)

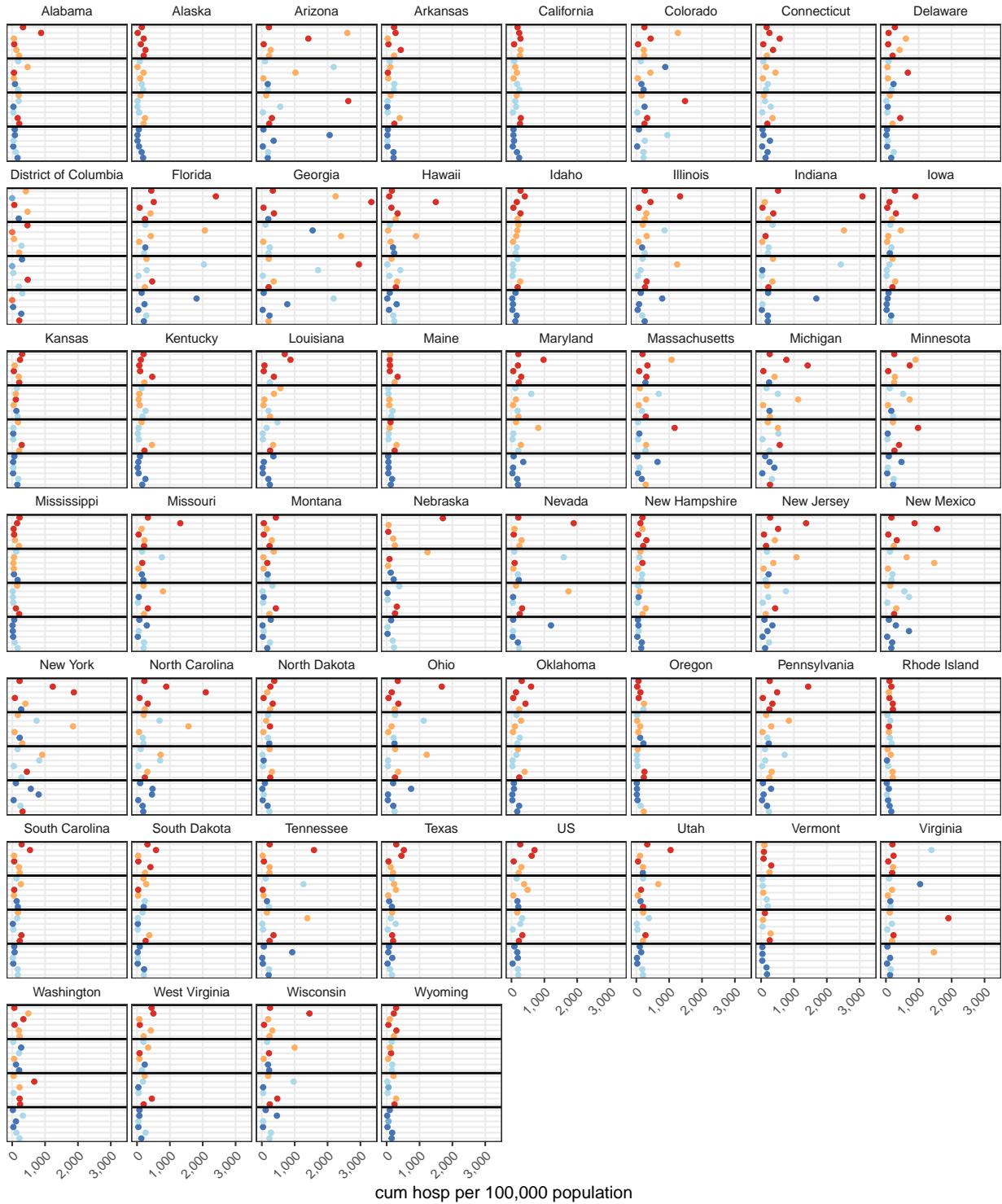


Fig. S21: COVID-19 Scenario Modeling Hub projections of cumulative hospitalizations for Round 16. Projections are collected from 6 models for four scenarios. Across locations (panels), projections are shown for each scenario (separated by horizontal black lines) from each model (individual points), points are colored based on the rank of that scenario for each model (blue is lowest projection, and red is highest projection). Data from [https://github.com/midas-network/covid19-scenario-hub\\_evaluation](https://github.com/midas-network/covid19-scenario-hub_evaluation), <https://zenodo.org/record/8415147>.

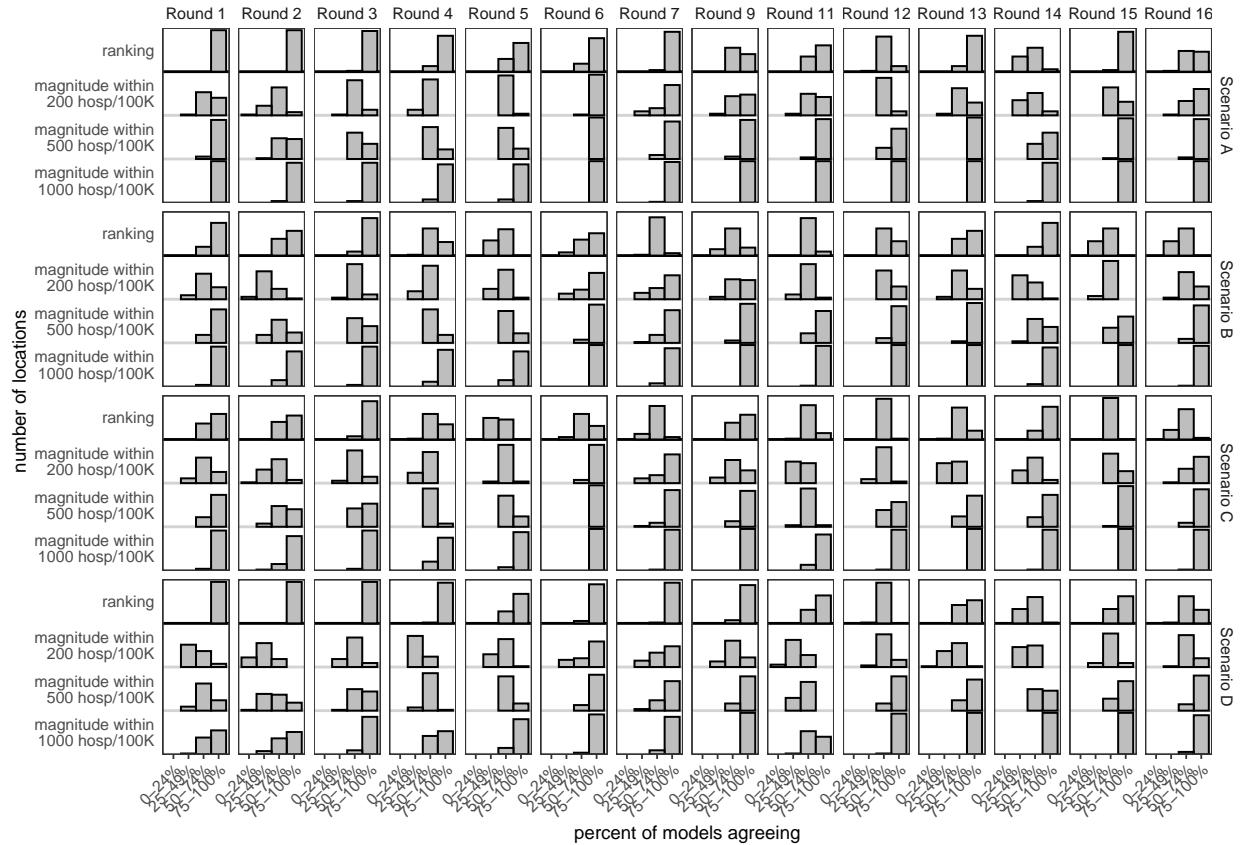


Fig. S22: Agreement of SMH projections across rounds (panel columns) and scenarios (panel rows). A single panel shows the maximum number of models that agree on different metrics, including agreeing about the rank of a particular scenario, and agreeing about the magnitude of the projection within a particular tolerance window. The tolerance window is defined based on a hospitalization per 100,000 population size for each location. The maximum number of agreeing models is summarized across locations, and the height of the bar shows the number of locations with a particular maximum number of agreeing models. Because a different number of modeling teams submitted to each round, the maximum agreement value on the xaxis varies. Data from agreement analysis of SMH projections.

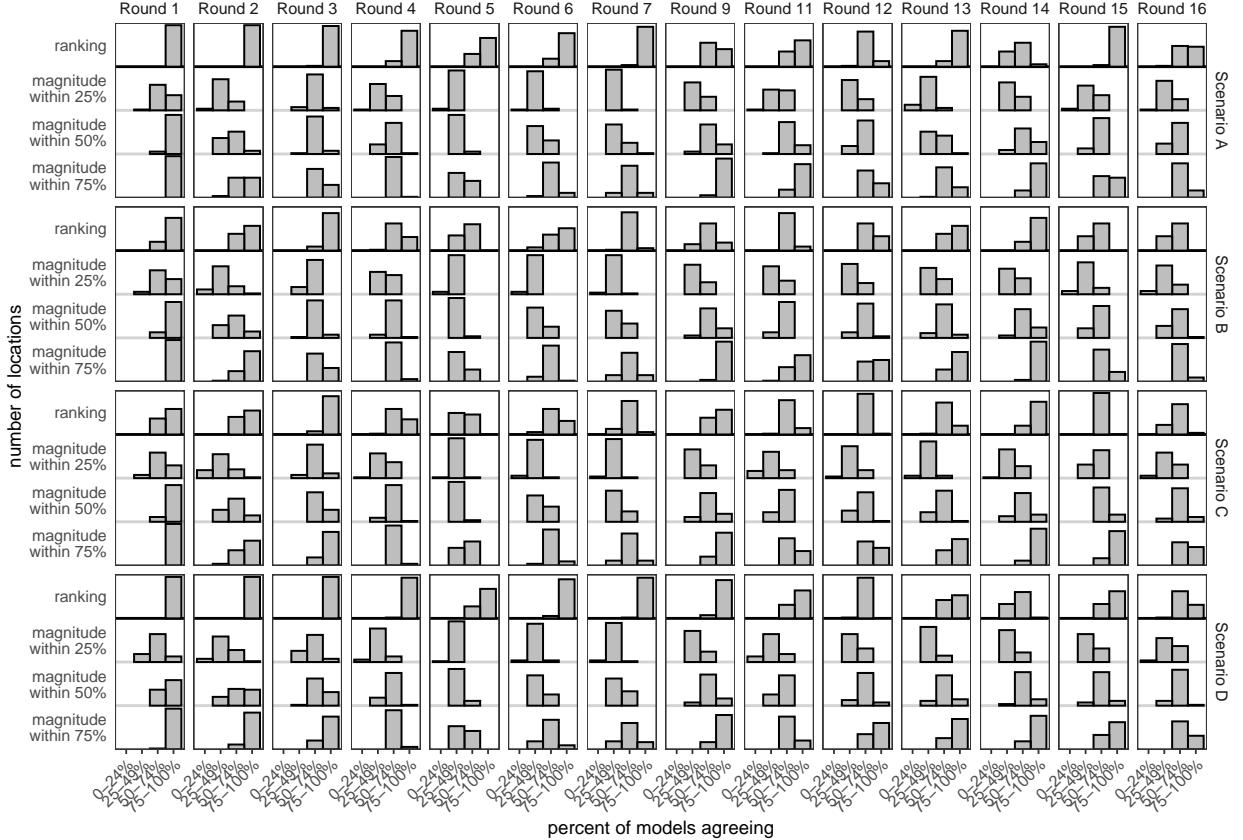


Fig. S23: Agreement of SMH projections across rounds (panel columns) and scenarios (panel rows). A single panel shows the maximum number of models that agree on different metrics, including agreeing about the rank of a particular scenario, and agreeing about the magnitude of the projection within a particular tolerance window. The tolerance window is defined relative to the projected magnitude; Fig. S24 shows the size of each window. The maximum number of agreeing models is summarized across locations, and the height of the bar shows the number of locations with a particular maximum number of agreeing models. Because a different number of modeling teams submitted to each round, the maximum agreement value on the xaxis varies. Data from agreement analysis of SMH projections.

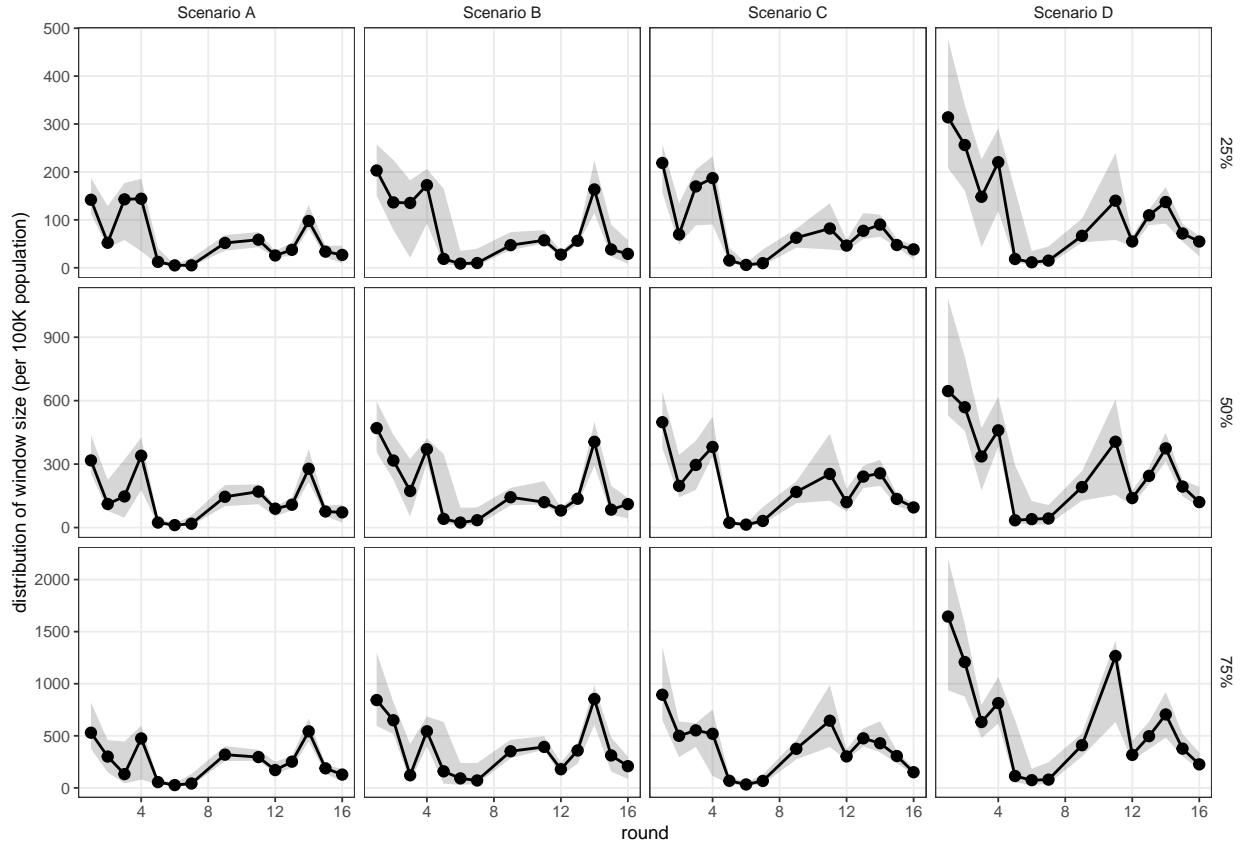


Fig. S24: The size of the window within which projections are considered to agree, for 25% of the projected magnitude, 50%, or  $\pm 75\%$ . The points show the average across all locations, and the ribbon show the 25th and 75th quantiles of this distribution. Data from SIR model simulations. Projected cumulative hospitalizations varies by round and scenario, and hence the window size (when defined relative to the projection) also varies. Data from agreement analysis of SMH projections.

### Round 1 (horizon: 26 weeks)

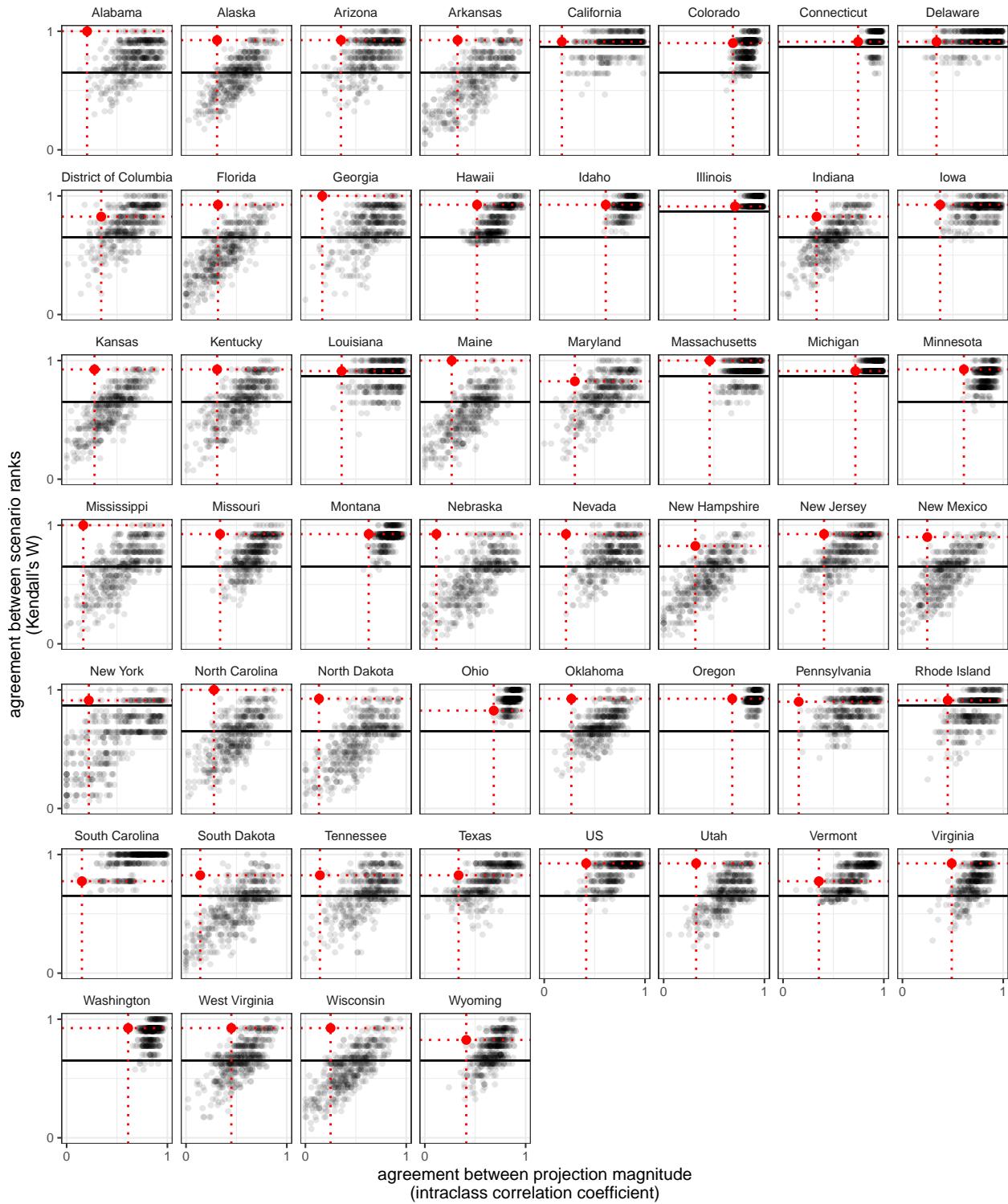


Fig. S25: Agreement for null model simulations and the corresponding SMH projections of cumulative hospitalizations for Round 1. Agreement is calculated for projection magnitude (as measured by intraclass correlation coefficient) and scenario ranking (as measured by Kendall's W). Across locations (panels), null model results are shown with black points and SMH results are shown with a red point. For simplicity, 500 out of the 1,000 null model simulations were randomly selected and plotted. The associated p-value for Kendall's W is shown with a solid horizontal line. Data from agreement analysis of SMH projections.

### Round 2 (horizon: 26 weeks)

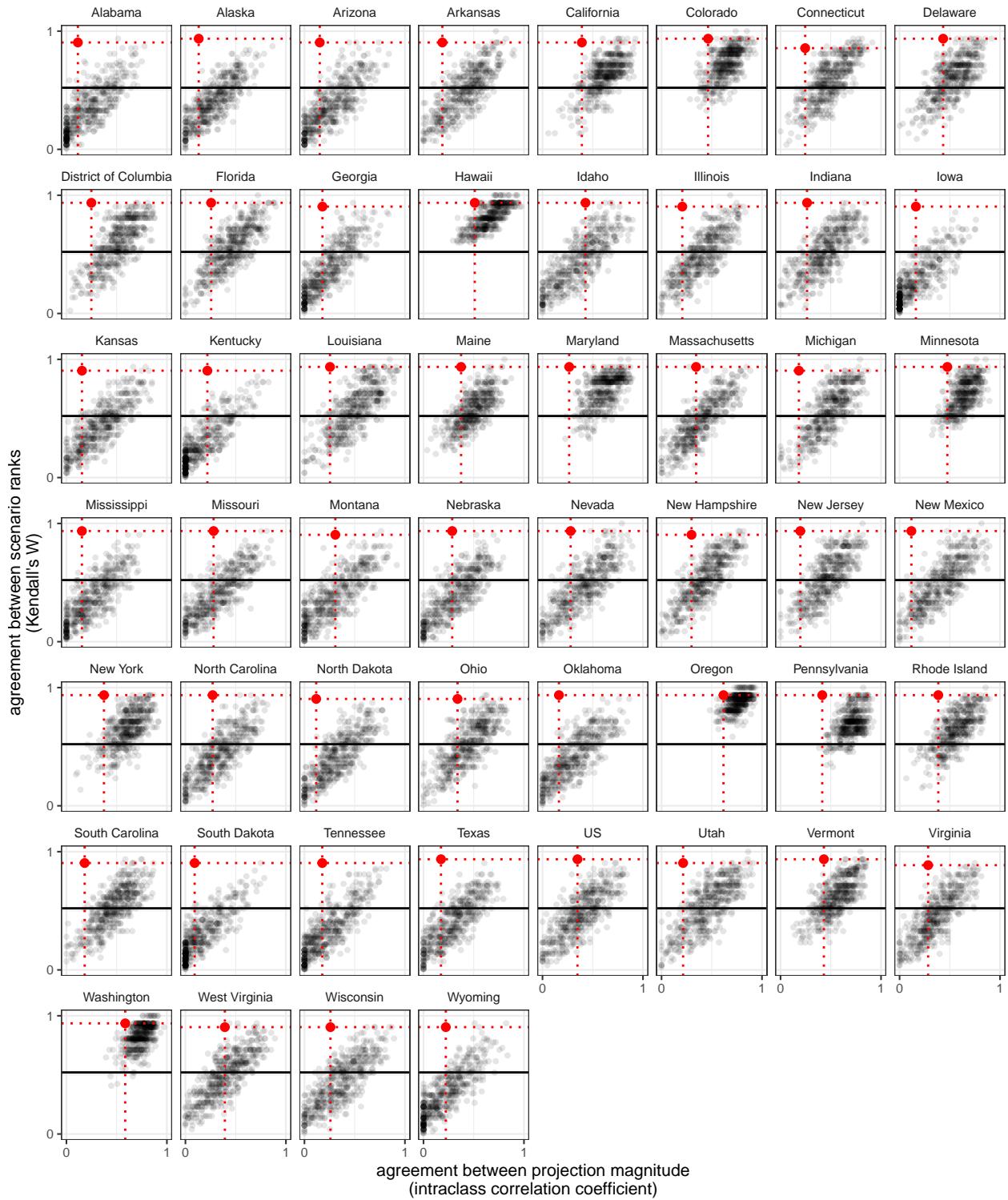


Fig. S26: Agreement for null model simulations and the corresponding SMH projections of cumulative hospitalizations for Round 2. Agreement is calculated for projection magnitude (as measured by intraclass correlation coefficient) and scenario ranking (as measured by Kendall's W). Across locations (panels), null model results are shown with black points and SMH results are shown with a red point. For simplicity, 500 out of the 1,000 null model simulations were randomly selected and plotted. The associated p-value for Kendall's W is shown with a solid horizontal line. Data from agreement analysis of SMH projections.

### Round 3 (horizon: 26 weeks)



Fig. S27: Agreement for null model simulations and the corresponding SMH projections of cumulative hospitalizations for Round 3. Agreement is calculated for projection magnitude (as measured by intraclass correlation coefficient) and scenario ranking (as measured by Kendall's W). Across locations (panels), null model results are shown with black points and SMH results are shown with a red point. For simplicity, 500 out of the 1,000 null model simulations were randomly selected and plotted. The associated p-value for Kendall's W is shown with a solid horizontal line. Data from agreement analysis of SMH projections.

### Round 4 (horizon: 26 weeks)

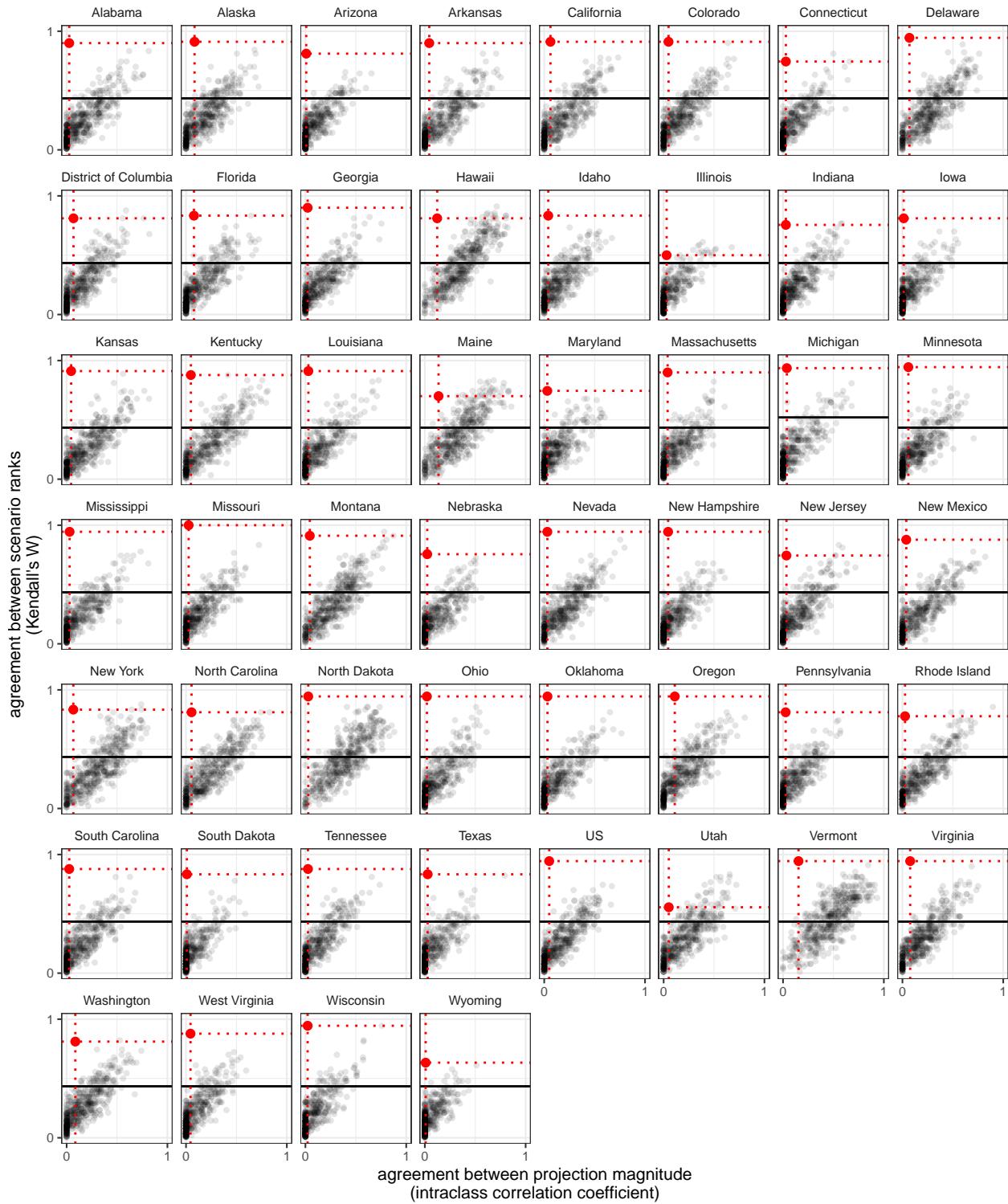


Fig. S28: Agreement for null model simulations and the corresponding SMH projections of cumulative hospitalizations for Round 4. Agreement is calculated for projection magnitude (as measured by intraclass correlation coefficient) and scenario ranking (as measured by Kendall's W). Across locations (panels), null model results are shown with black points and SMH results are shown with a red point. For simplicity, 500 out of the 1,000 null model simulations were randomly selected and plotted. The associated p-value for Kendall's W is shown with a solid horizontal line. Data from agreement analysis of SMH projections.

### Round 5 (horizon: 26 weeks)

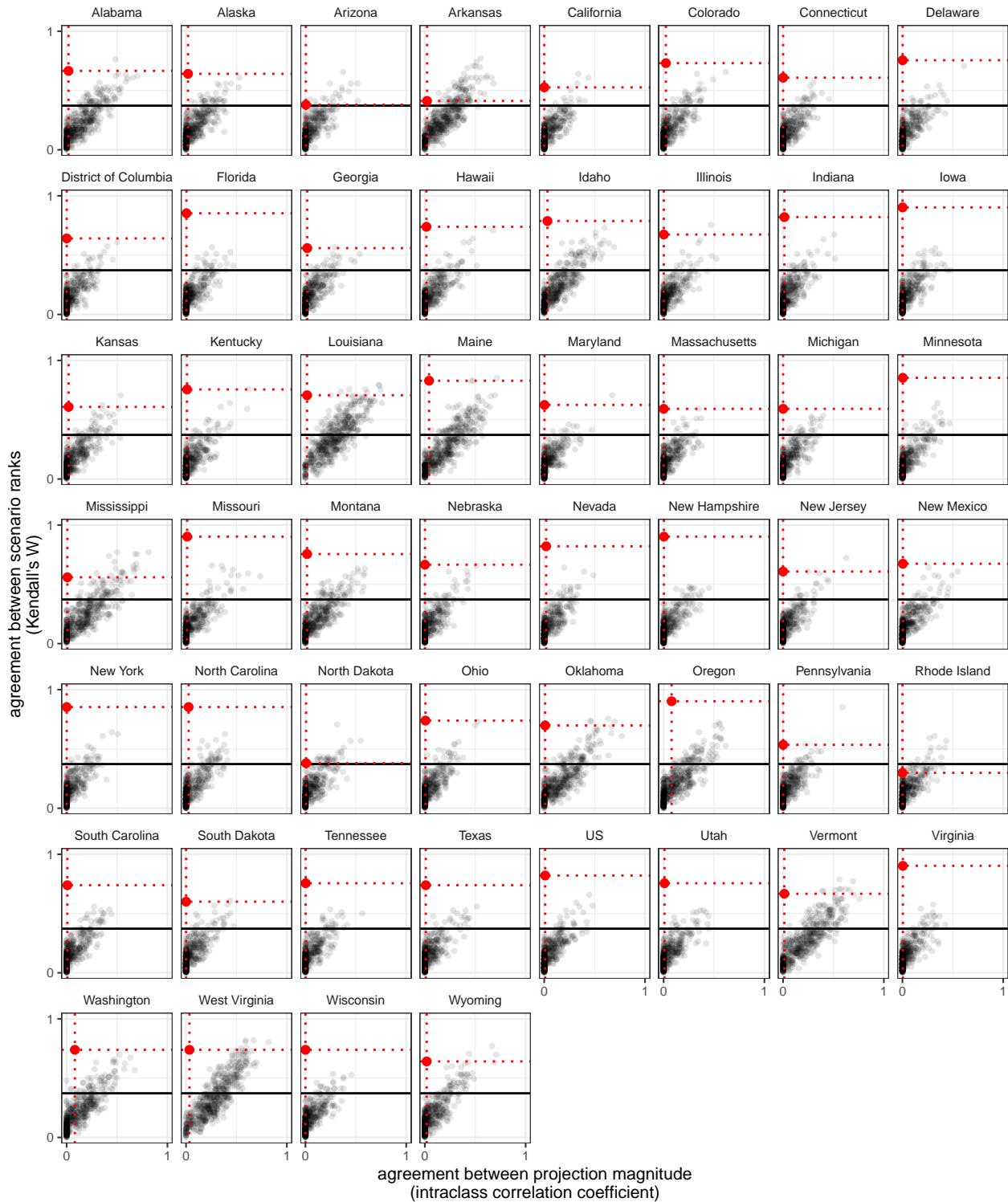


Fig. S29: Agreement for null model simulations and the corresponding SMH projections of cumulative hospitalizations for Round 5. Agreement is calculated for projection magnitude (as measured by intraclass correlation coefficient) and scenario ranking (as measured by Kendall's W). Across locations (panels), null model results are shown with black points and SMH results are shown with a red point. For simplicity, 500 out of the 1,000 null model simulations were randomly selected and plotted. The associated p-value for Kendall's W is shown with a solid horizontal line. Data from agreement analysis of SMH projections.

### Round 6 (horizon: 26 weeks)

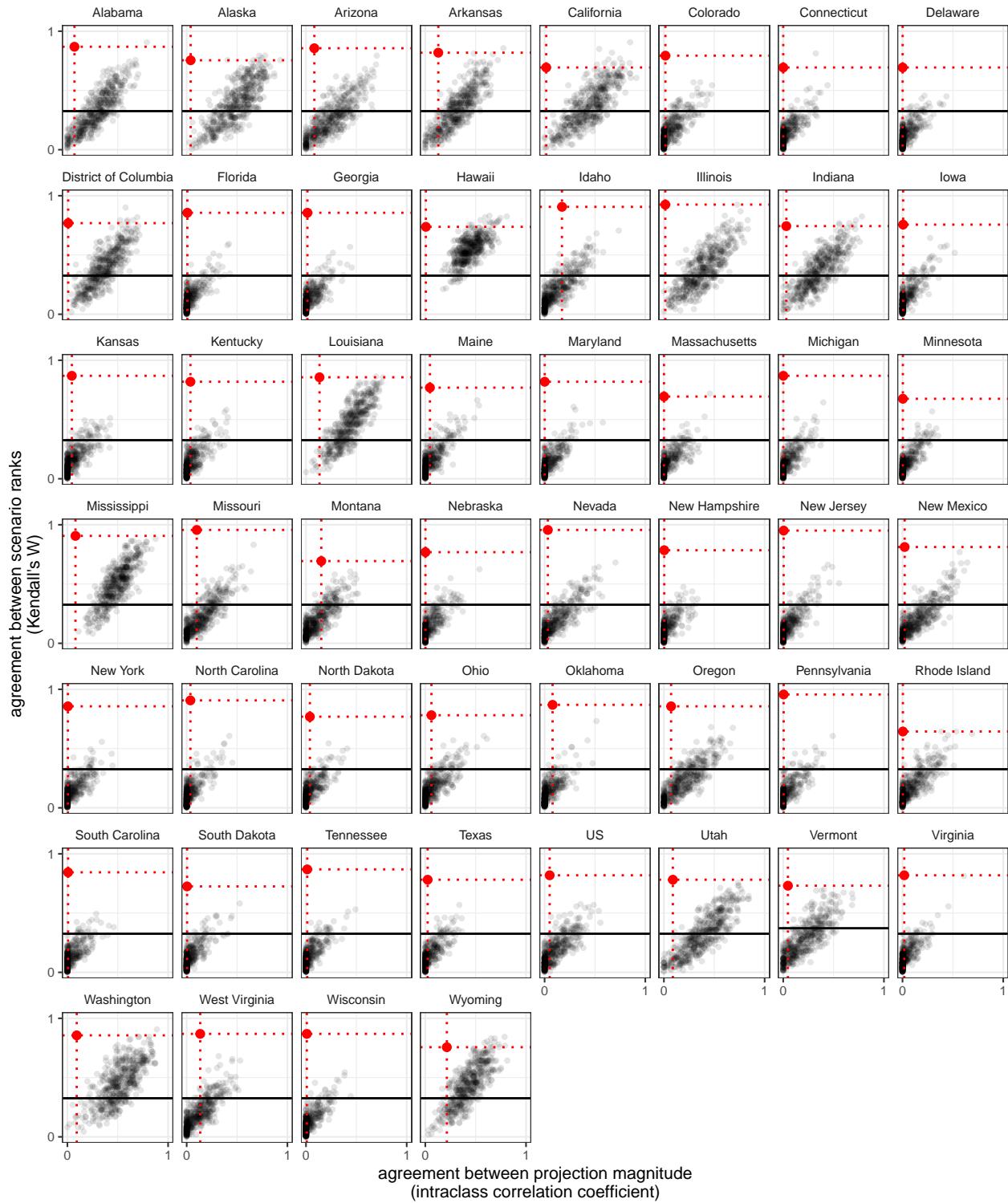


Fig. S30: Agreement for null model simulations and the corresponding SMH projections of cumulative hospitalizations for Round 6. Agreement is calculated for projection magnitude (as measured by intraclass correlation coefficient) and scenario ranking (as measured by Kendall's W). Across locations (panels), null model results are shown with black points and SMH results are shown with a red point. For simplicity, 500 out of the 1,000 null model simulations were randomly selected and plotted. The associated p-value for Kendall's W is shown with a solid horizontal line. Data from agreement analysis of SMH projections.

### Round 7 (horizon: 26 weeks)

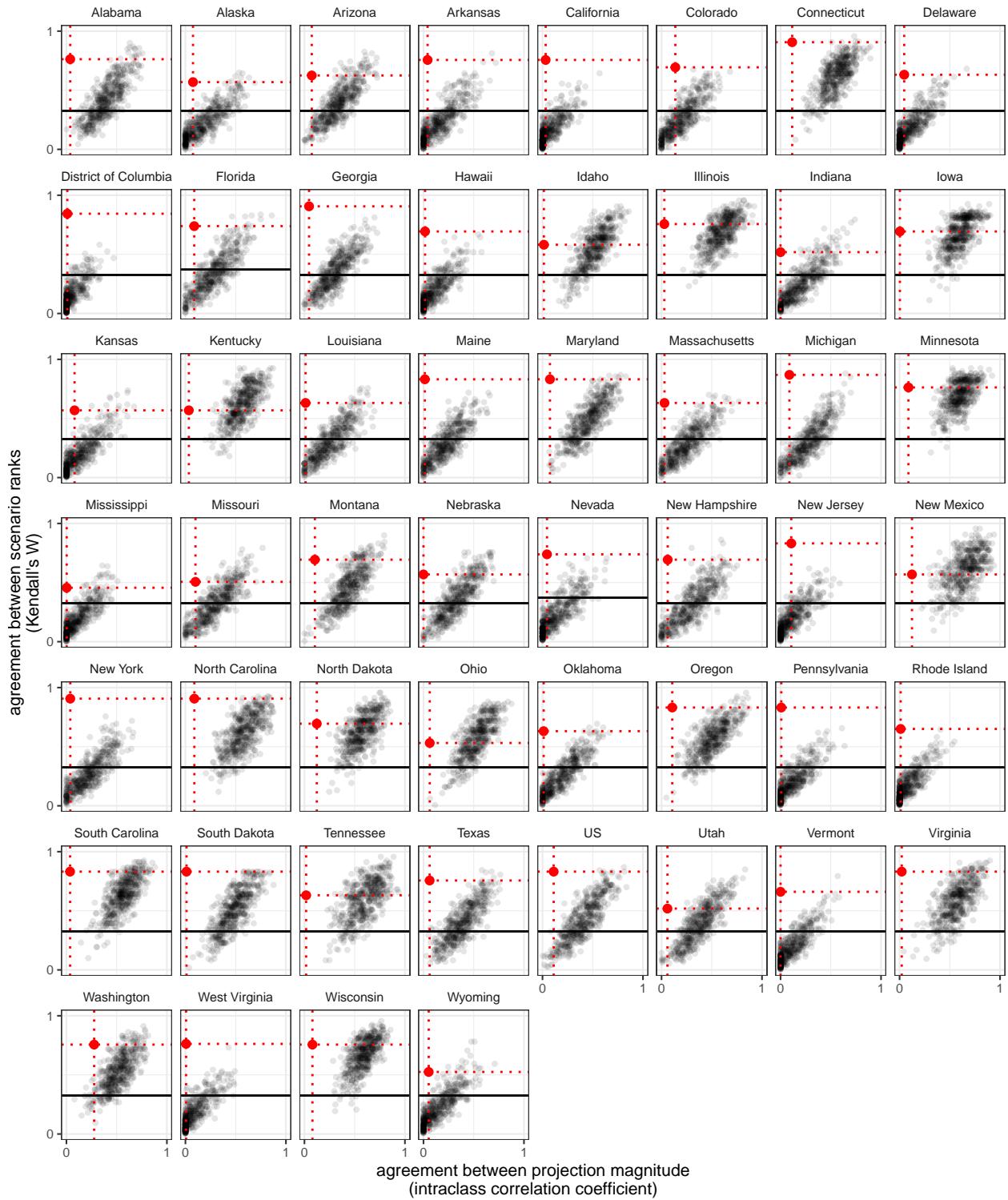


Fig. S31: Agreement for null model simulations and the corresponding SMH projections of cumulative hospitalizations for Round 7. Agreement is calculated for projection magnitude (as measured by intraclass correlation coefficient) and scenario ranking (as measured by Kendall's W). Across locations (panels), null model results are shown with black points and SMH results are shown with a red point. For simplicity, 500 out of the 1,000 null model simulations were randomly selected and plotted. The associated p-value for Kendall's W is shown with a solid horizontal line. Data from agreement analysis of SMH projections.

### Round 9 (horizon: 26 weeks)

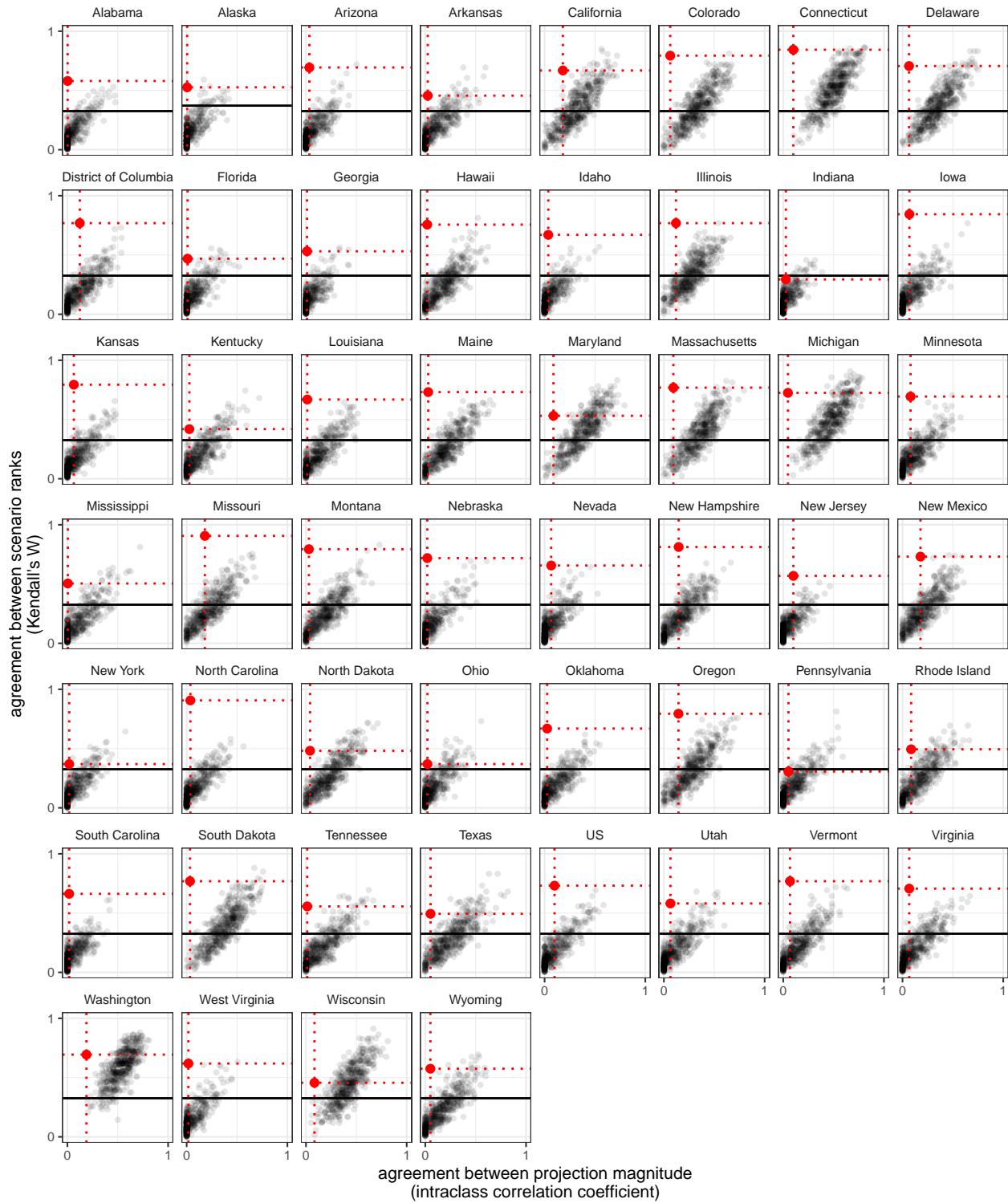


Fig. S32: Agreement for null model simulations and the corresponding SMH projections of cumulative hospitalizations for Round 9. Agreement is calculated for projection magnitude (as measured by intraclass correlation coefficient) and scenario ranking (as measured by Kendall's W). Across locations (panels), null model results are shown with black points and SMH results are shown with a red point. For simplicity, 500 out of the 1,000 null model simulations were randomly selected and plotted. The associated p-value for Kendall's W is shown with a solid horizontal line. Data from agreement analysis of SMH projections.

### Round 11 (horizon: 12 weeks)

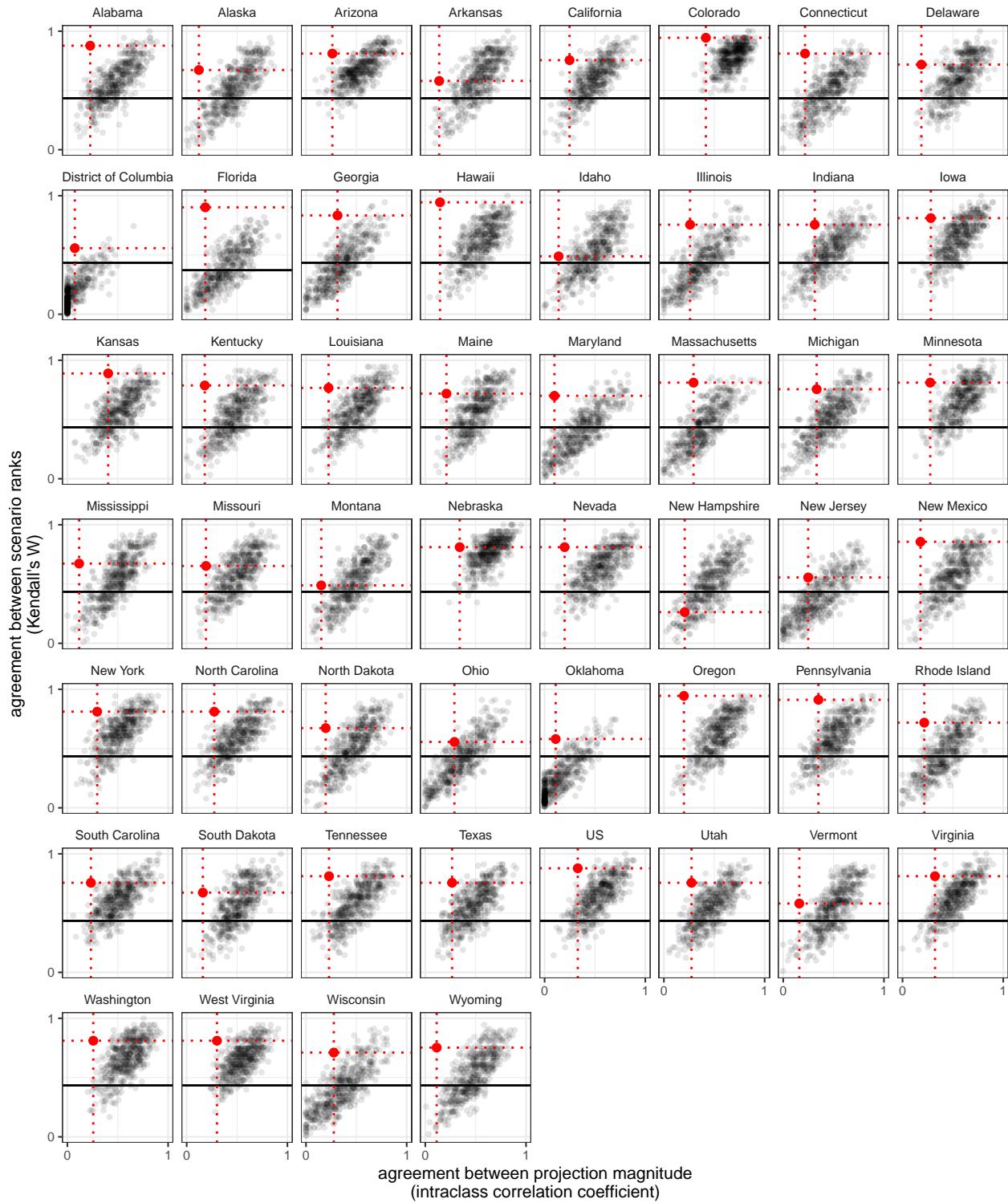


Fig. S33: Agreement for null model simulations and the corresponding SMH projections of cumulative hospitalizations for Round 11. Agreement is calculated for projection magnitude (as measured by intraclass correlation coefficient) and scenario ranking (as measured by Kendall's W). Across locations (panels), null model results are shown with black points and SMH results are shown with a red point. For simplicity, 500 out of the 1,000 null model simulations were randomly selected and plotted. The associated p-value for Kendall's W is shown with a solid horizontal line. Data from agreement analysis of SMH projections.

### Round 12 (horizon: 12 weeks)

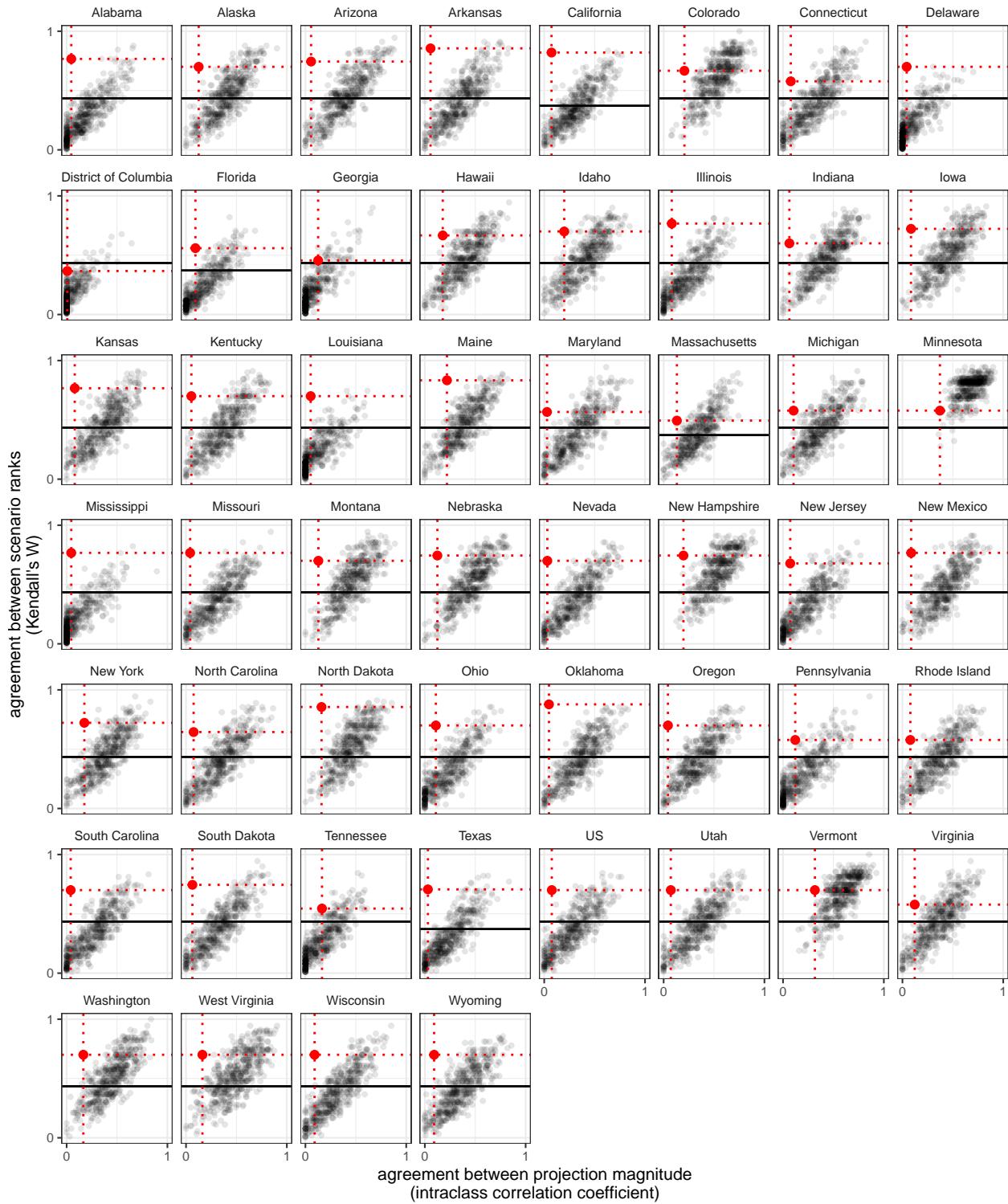


Fig. S34: Agreement for null model simulations and the corresponding SMH projections of cumulative hospitalizations for Round 12. Agreement is calculated for projection magnitude (as measured by intraclass correlation coefficient) and scenario ranking (as measured by Kendall's W). Across locations (panels), null model results are shown with black points and SMH results are shown with a red point. For simplicity, 500 out of the 1,000 null model simulations were randomly selected and plotted. The associated p-value for Kendall's W is shown with a solid horizontal line. Data from agreement analysis of SMH projections.

### Round 13 (horizon: 52 weeks)

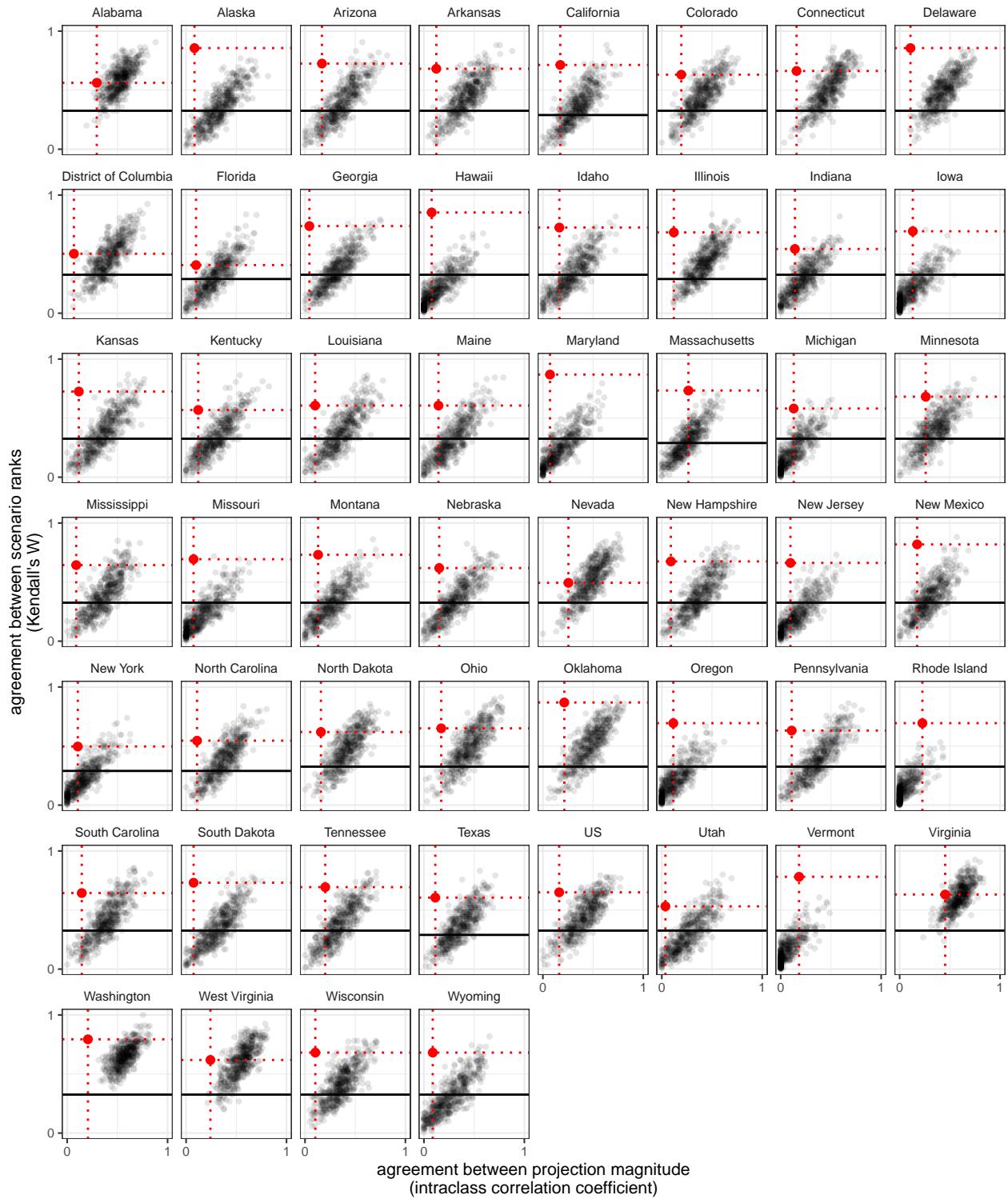


Fig. S35: Agreement for null model simulations and the corresponding SMH projections of cumulative hospitalizations for Round 13. Agreement is calculated for projection magnitude (as measured by intraclass correlation coefficient) and scenario ranking (as measured by Kendall's W). Across locations (panels), null model results are shown with black points and SMH results are shown with a red point. For simplicity, 500 out of the 1,000 null model simulations were randomly selected and plotted. The associated p-value for Kendall's W is shown with a solid horizontal line. Data from agreement analysis of SMH projections.

### Round 14 (horizon: 50 weeks)

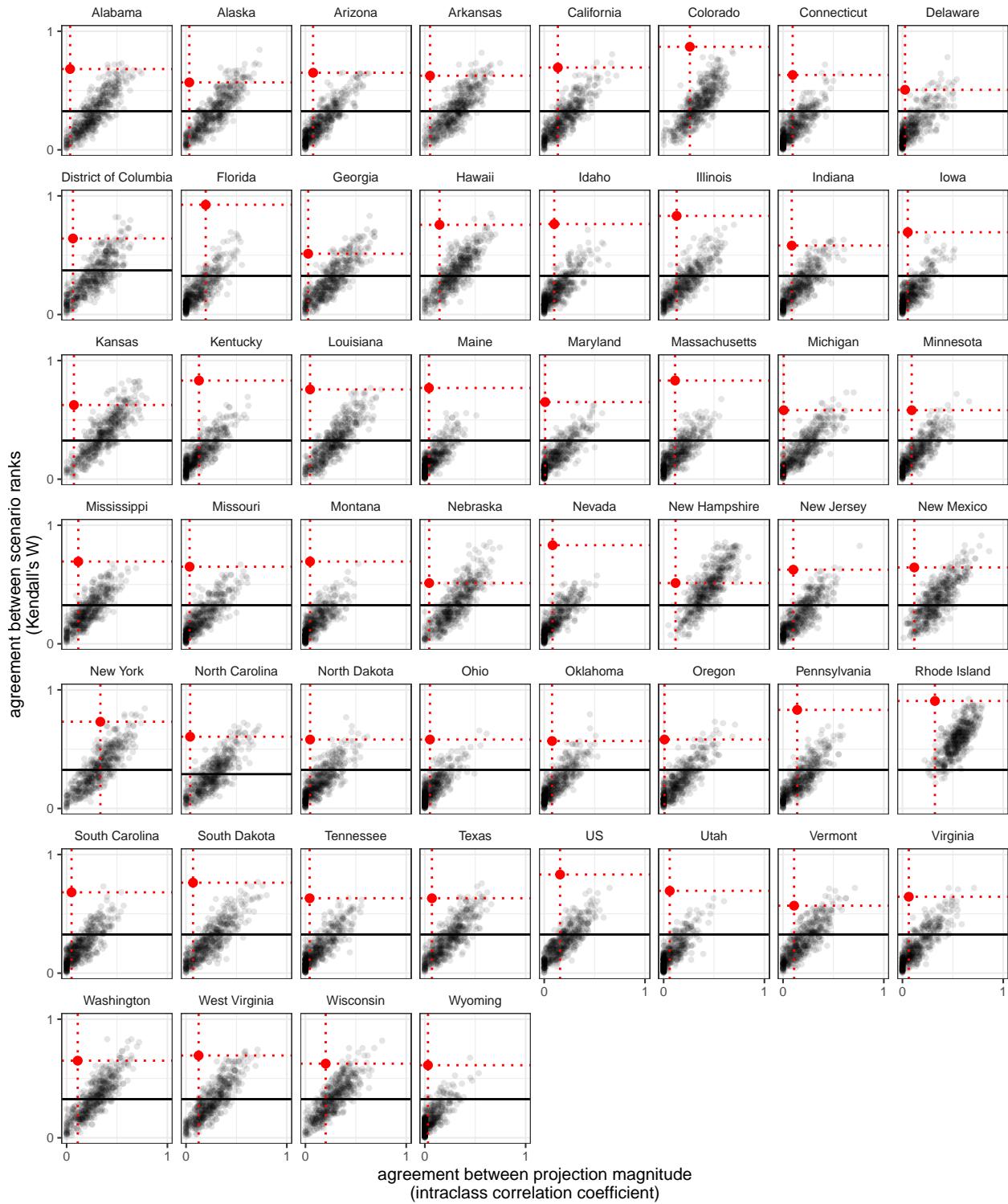


Fig. S36: Agreement for null model simulations and the corresponding SMH projections of cumulative hospitalizations for Round 14. Agreement is calculated for projection magnitude (as measured by intraclass correlation coefficient) and scenario ranking (as measured by Kendall's W). Across locations (panels), null model results are shown with black points and SMH results are shown with a red point. For simplicity, 500 out of the 1,000 null model simulations were randomly selected and plotted. The associated p-value for Kendall's W is shown with a solid horizontal line. Data from agreement analysis of SMH projections.

### Round 15 (horizon: 40 weeks)

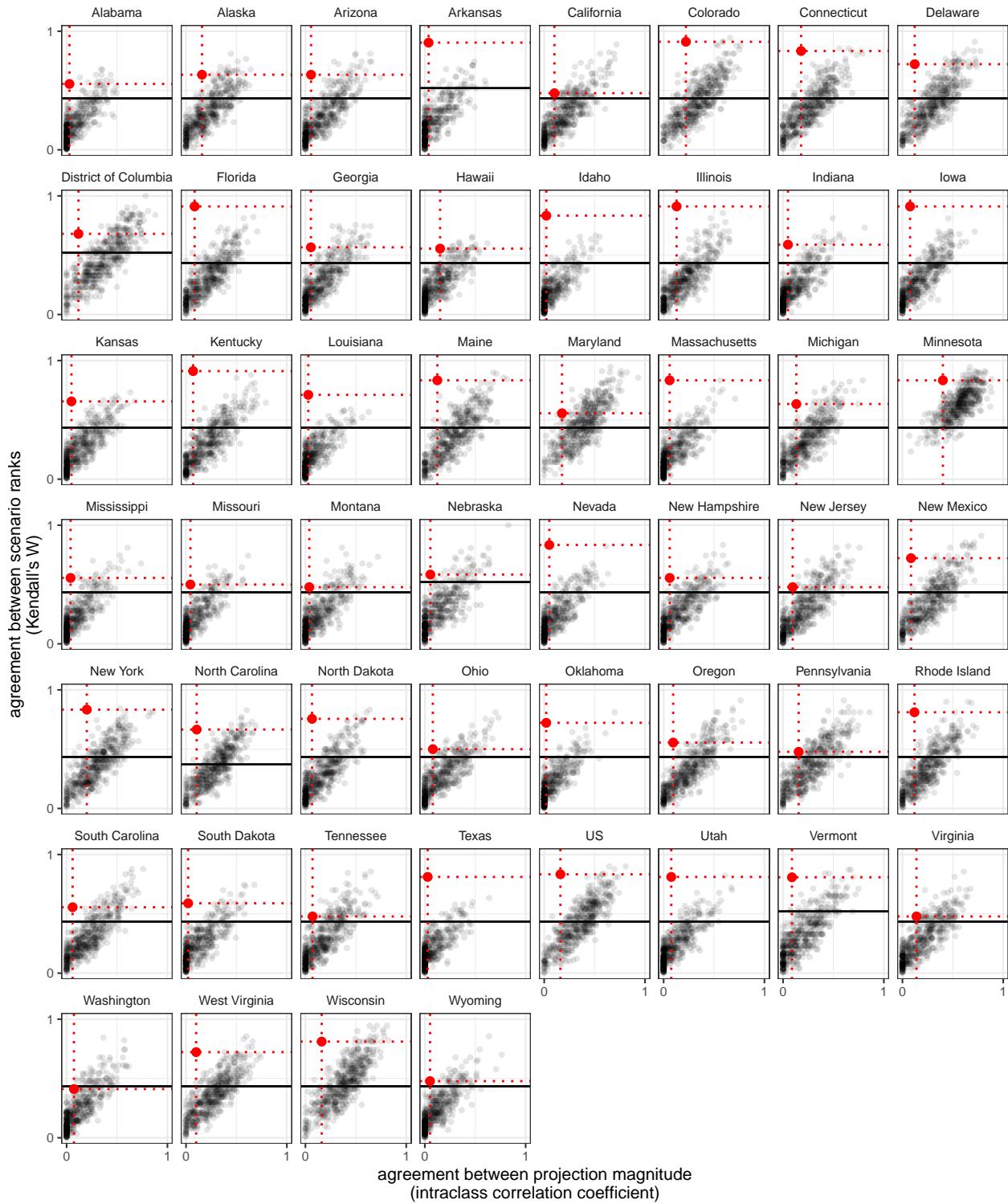


Fig. S37: Agreement for null model simulations and the corresponding SMH projections of cumulative hospitalizations for Round 15. Agreement is calculated for projection magnitude (as measured by intraclass correlation coefficient) and scenario ranking (as measured by Kendall's W). Across locations (panels), null model results are shown with black points and SMH results are shown with a red point. For simplicity, 500 out of the 1,000 null model simulations were randomly selected and plotted. The associated p-value for Kendall's W is shown with a solid horizontal line. Data from agreement analysis of SMH projections.

### Round 16 (horizon: 26 weeks)

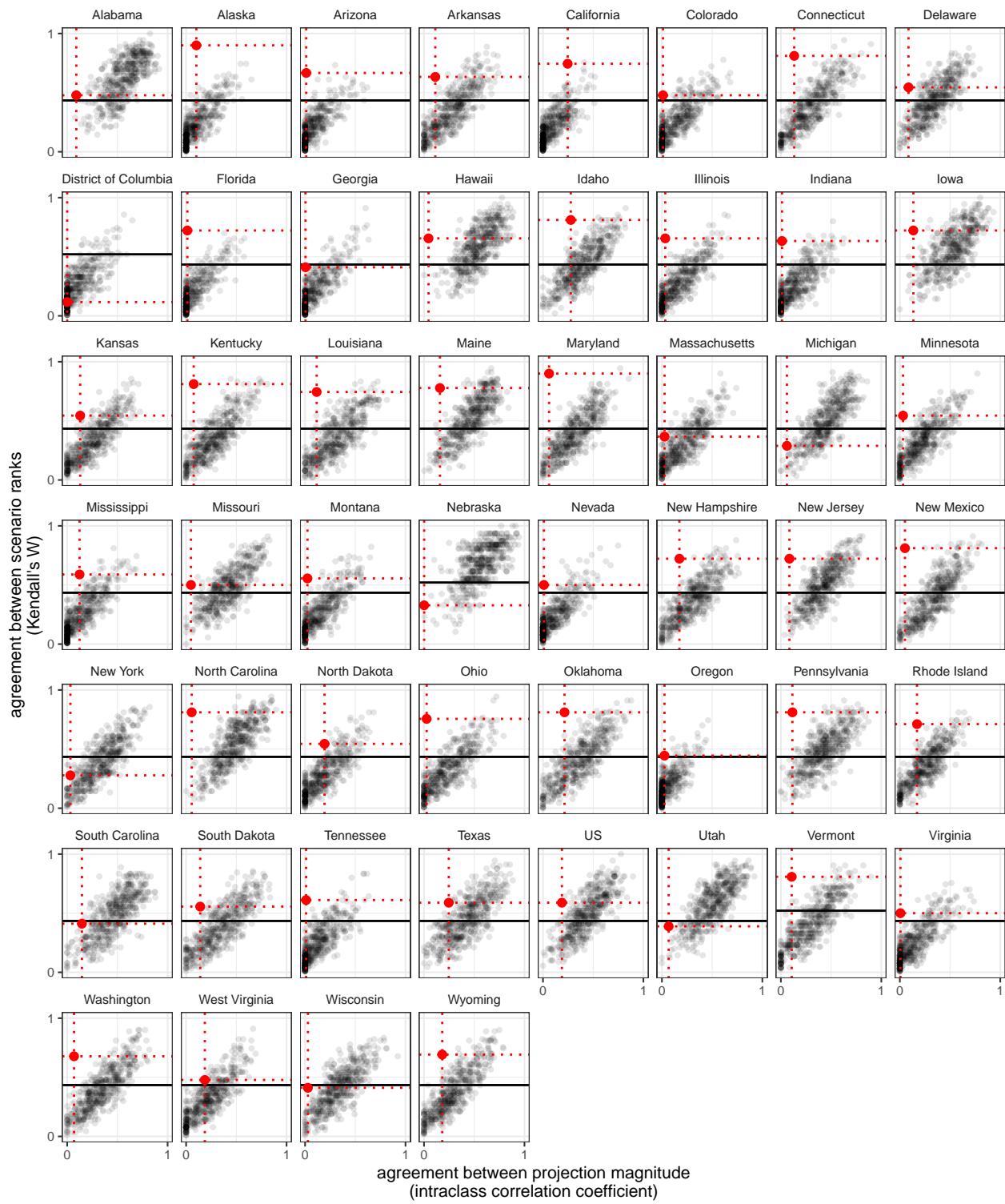


Fig. S38: Agreement for null model simulations and the corresponding SMH projections of cumulative hospitalizations for Round 16. Agreement is calculated for projection magnitude (as measured by intraclass correlation coefficient) and scenario ranking (as measured by Kendall's W). Across locations (panels), null model results are shown with black points and SMH results are shown with a red point. For simplicity, 500 out of the 1,000 null model simulations were randomly selected and plotted. The associated p-value for Kendall's W is shown with a solid horizontal line. Data from agreement analysis of SMH projections.

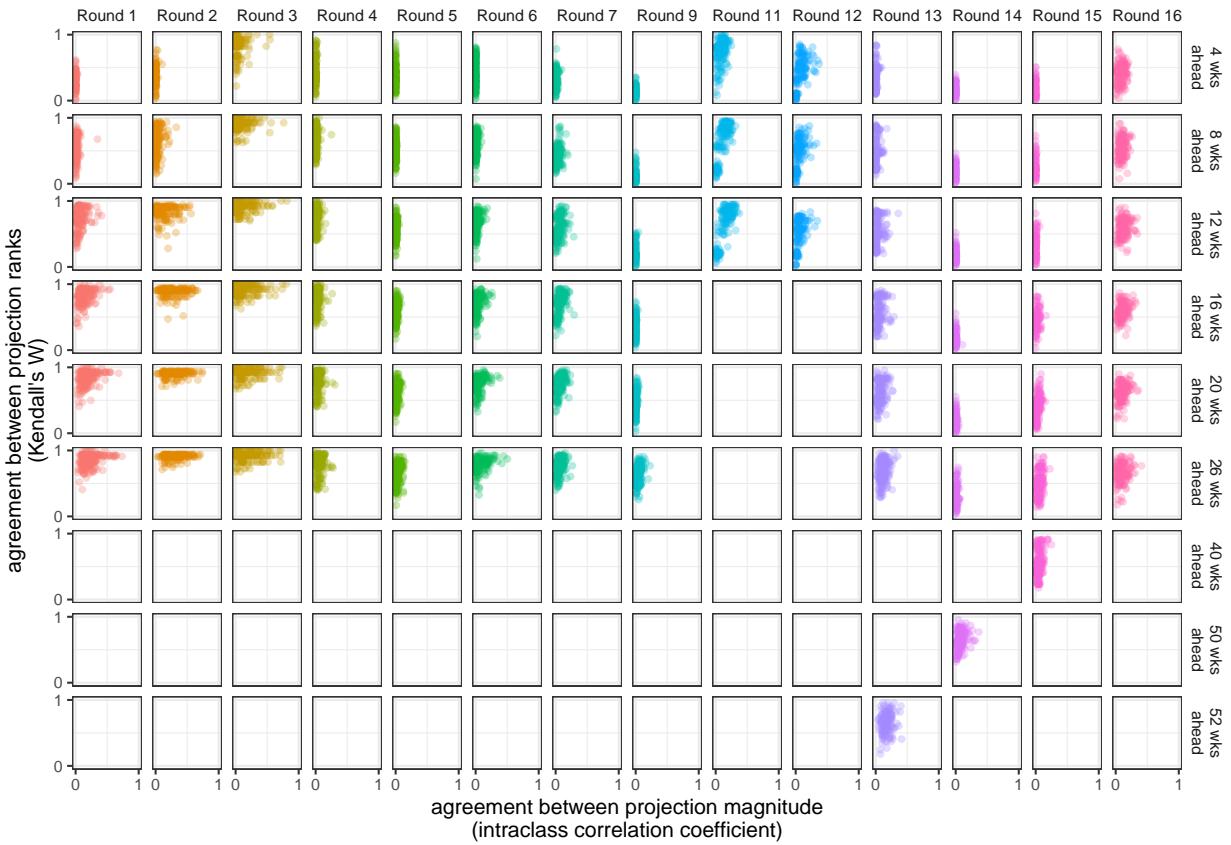


Fig. S39: Agreement between projection magnitude and ranks across evaluated SMH rounds (columns) and SMH projection horizons (rows) for SMH Q25. Agreement between projection magnitude is measured by intraclass correlation coefficient, and agreement between scenario ranks is measured by Kendall's W. Each point represents results for a set of projections from a single round, target, location, week. Maximum projection horizon is shown for each round, but horizon varies by round so not all horizons are available for all rounds. Data from agreement analysis of SMH projections.

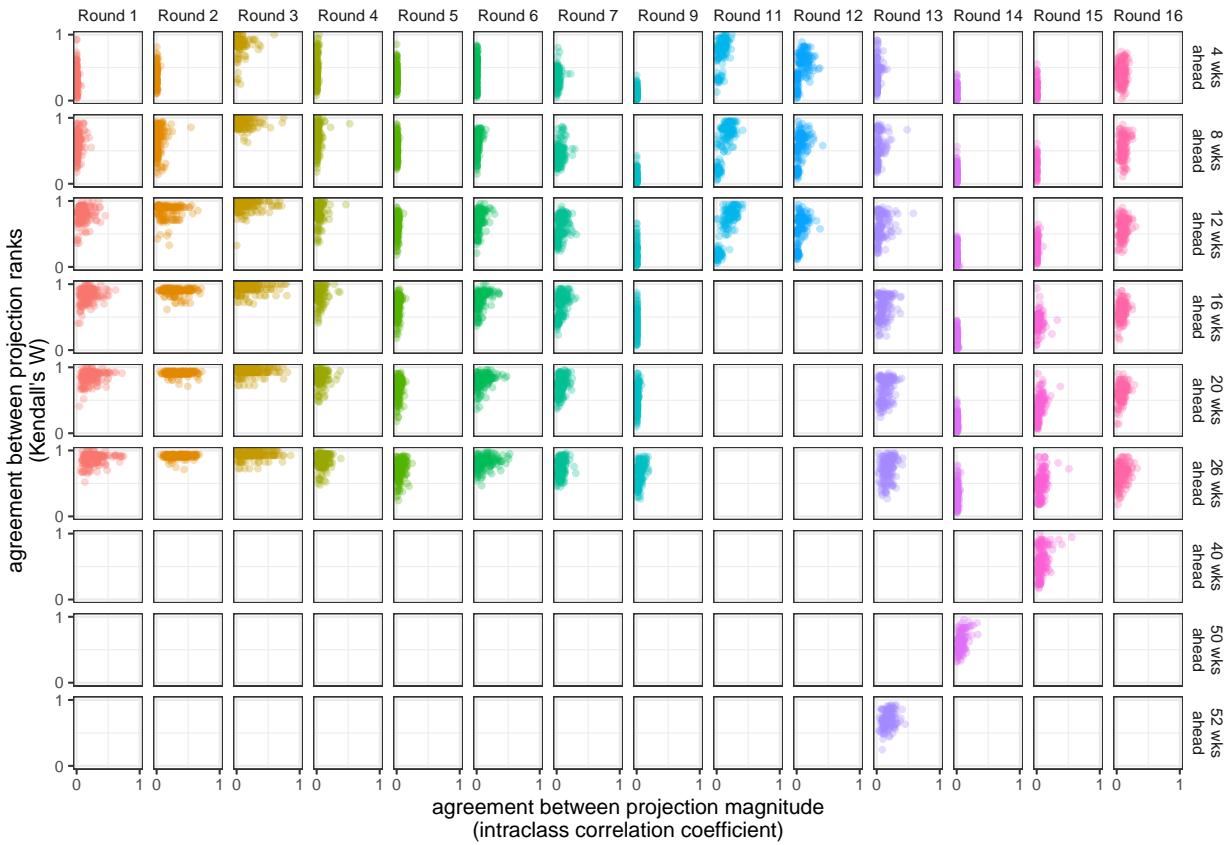


Fig. S40: Agreement between projection magnitude and ranks across evaluated SMH rounds (columns) and SMH projection horizons (rows) for SMH Q50. Agreement between projection magnitude is measured by intraclass correlation coefficient, and agreement between scenario ranks is measured by Kendall's W. Each point represents results for a set of projections from a single round, target, location, week. Maximum projection horizon is shown for each round, but horizon varies by round so not all horizons are available for all rounds. Data from agreement analysis of SMH projections.

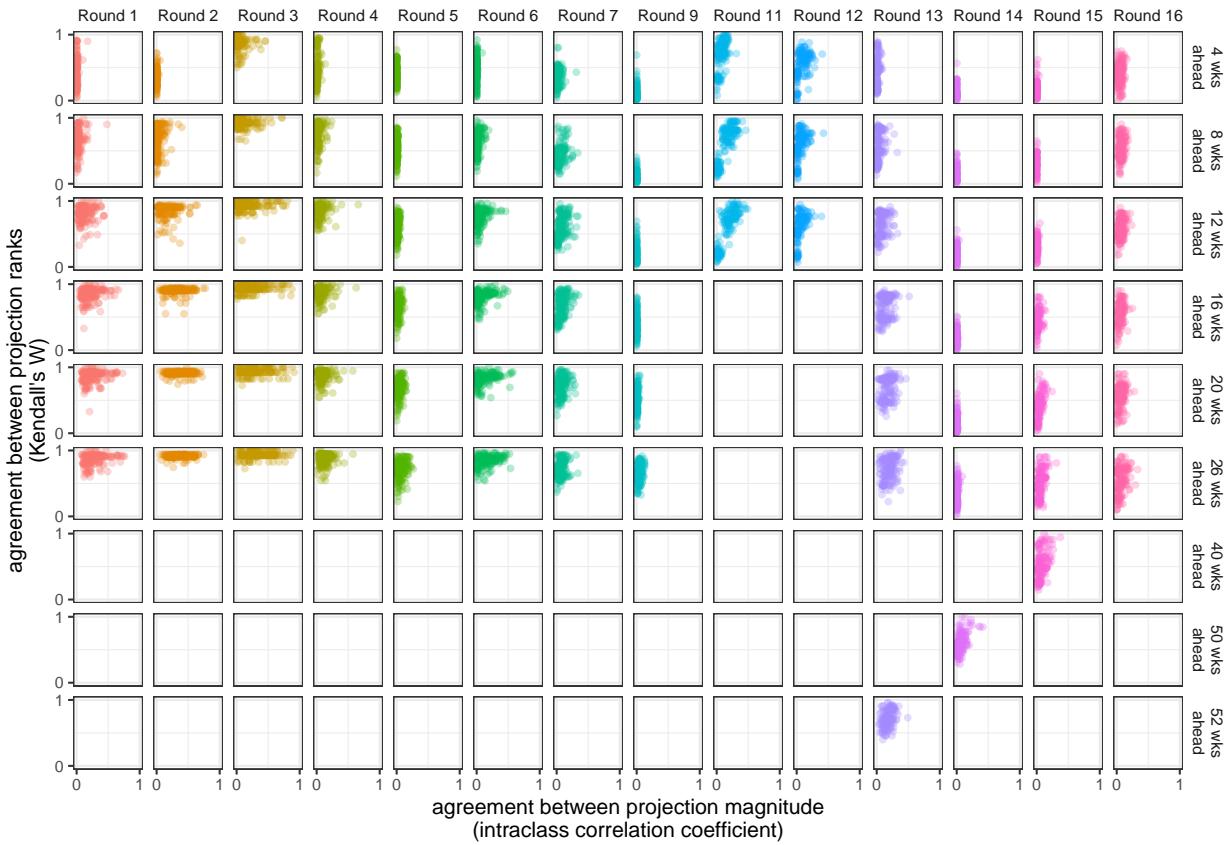


Fig. S41: Agreement between projection magnitude and ranks across evaluated SMH rounds (columns) and SMH projection horizons (rows) for SMH Q75. Agreement between projection magnitude is measured by intraclass correlation coefficient, and agreement between scenario ranks is measured by Kendall's W. Each point represents results for a set of projections from a single round, target, location, week. Maximum projection horizon is shown for each round, but horizon varies by round so not all horizons are available for all rounds. Data from agreement analysis of SMH projections.

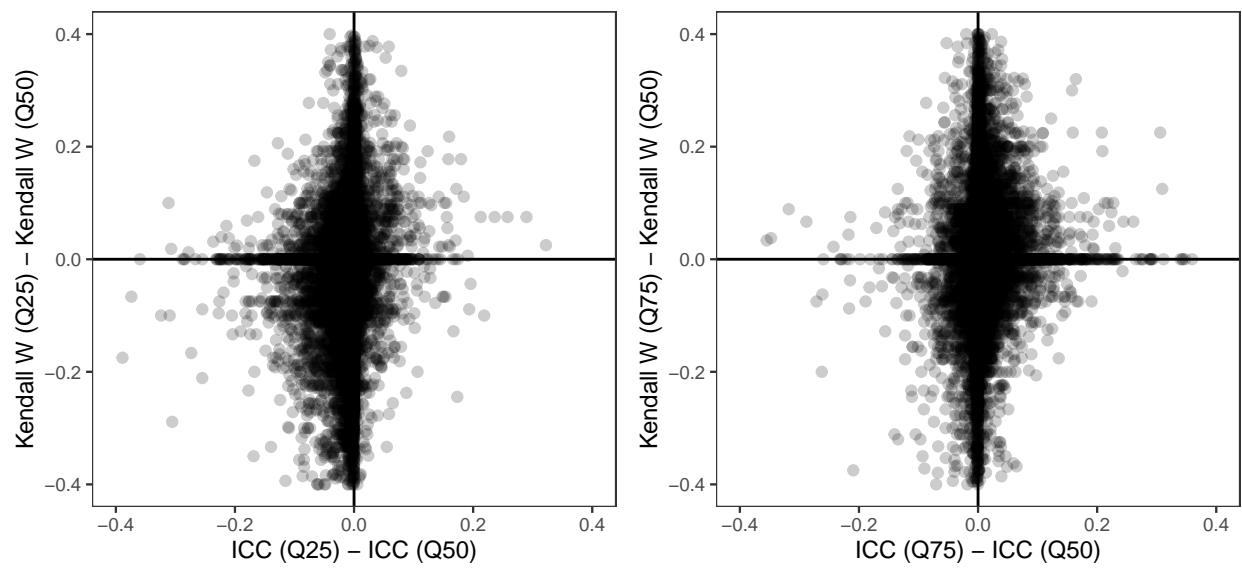


Fig. S42: Comparison of intraclass correlation coefficient (ICC) and Kendall's W between quantiles. Each metric was assessed on the projected median, as well as Q25 and Q75. The left panel shows comparisons between Q25 and Q50, and the right panel shows comparisons between Q75 and Q50. Each point represents one projection (for a single round, location). Data from agreement analysis of SMH projections.