Teodor Fredriksson
Causality and Bayesian Data Analysis
August 24, 2023

Bayesian Data Analysis

In this section we present our results and assess the model's validity. We assess the MCMC simulations and perform Posterior Predictive Checks to see that the posterior simulates the replications correctly. We follow the guidelines by [1, 2] for choosing prior, asses MCMC and performing posterior predictive checks.

Assesing the MCMC chains

We utilize the `cmdstanr` package in the `R` programming language and then fit the model using four parallel chains using 200 warmup iterations and an additional 2000 iterations for a total of 2200 iterations. We received no divergent iterations during the simulation, so the MCMC simulation cannot be considered invalid. The trace plots of the chains are found in Figures 1, 2, 3, 4, 5 and 6. The purpose of these train plots is to diagnose the MCMC simulation. The chains should be mixed without any patterns [3], which is the case here. Tables 1, 2, 3, 4, 5 and 6 contains diagnostics data of the posterior estimates. $\hat{R}$ is the Gelman-Rubin Potential Scale Reduction and should preferably be $\hat{R} < 1.05$ [3] and in our case $\hat{R} \approx 1.00$ so we are satisfied. The ess_ bulk and ess_tail values estimate the number of efficient samples. As a rule of thumb, at least 200 efficient samples are required. In our case, we have over 416 efficient samples, so we are satisfied.

Table 1. Ranking of the algorithms

Variable	Rhat	ess_bulk	ess_tail
a_alg[1]	1.023	424.934	455.527
a_alg[2]	1.023	422.580	431.206
a_alg[3]	1.022	441.915	451.605
a_alg[4]	1.023	424.820	490.687
a_alg[5]	1.023	427.856	466.328
a_alg[6]	1.023	422.975	434.168
a_alg[7]	1.022	427.910	434.017
a_alg[8]	1.023	426.002	460.934
a_alg[9]	1.022	427.561	430.549
a_alg[10]	1.022	426.692	451.952
a_alg[11]	1.023	416.875	430.375
a_alg[12]	1.023	419.819	439.080
a_alg[13]	1.023	420.713	455.159
a_alg[14]	1.023	422.586	425.631

Posterior Predictive Checks

Figures 7, 8, 9, 10, 11, and 12, shows the posterior distribution of the estimated parameters for each scenario, aggregated, 10%, 25%, 50%, 75% and 90%.

Figure 13, 14, 15, 16, 17 and 18 illustrates the distributions of replicated datasets drawn from the posterior distribution. It is clear from the figures that models can simulate data that is similar to the observed data. Figure 19, 20, 21, 22, 23 and 30 shows how well the posterior predictive distribution captures the mean. According to the Figures the model captures the mean well. Similarly, Figures 25, 26, 27, 28, 29, 30 shows how well the mean is captured by the posterior predictive distribution of each algorithm separately. The figures shows that the mean is well captured for all algorithms except for ("agg") peikonal, modularity mbo, ("75%") peikonal, plaplace, poisson mbo, ("90%") plaplace

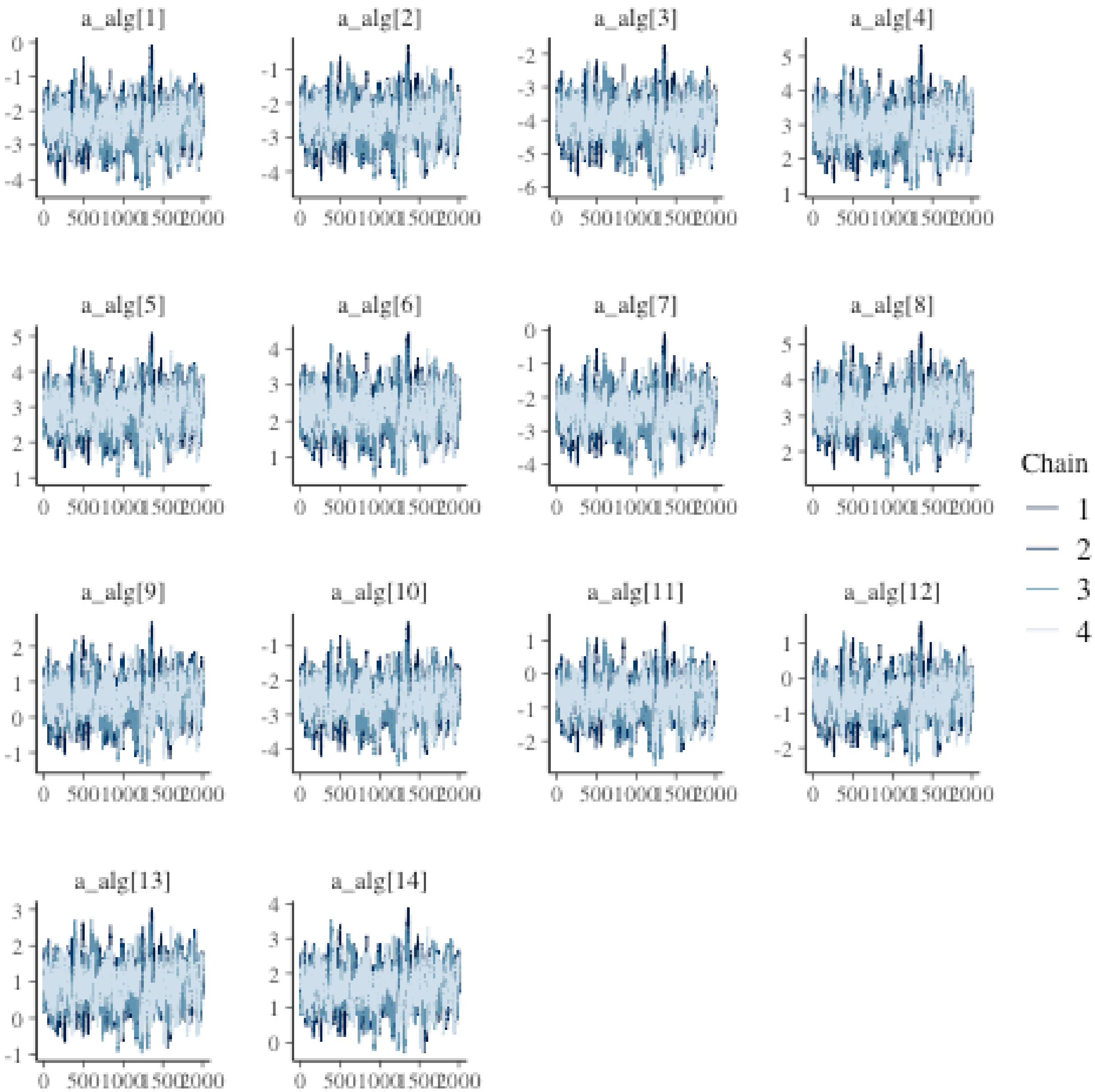

FIGURE 1. Simulated MCMC chains for the strength parameters a_i

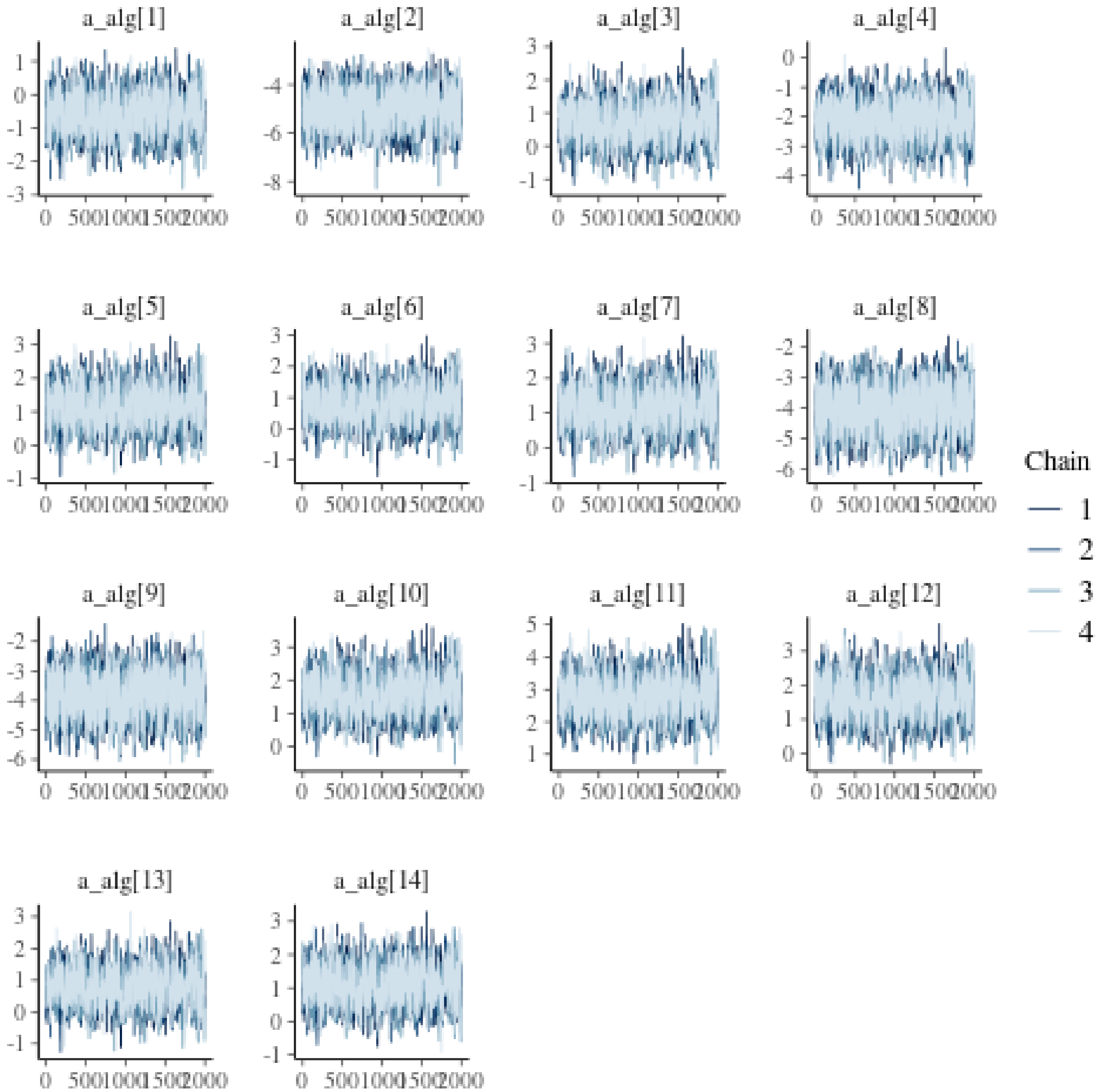

FIGURE 2. Simulated MCMC chains for the strength parameters a_i (10%)

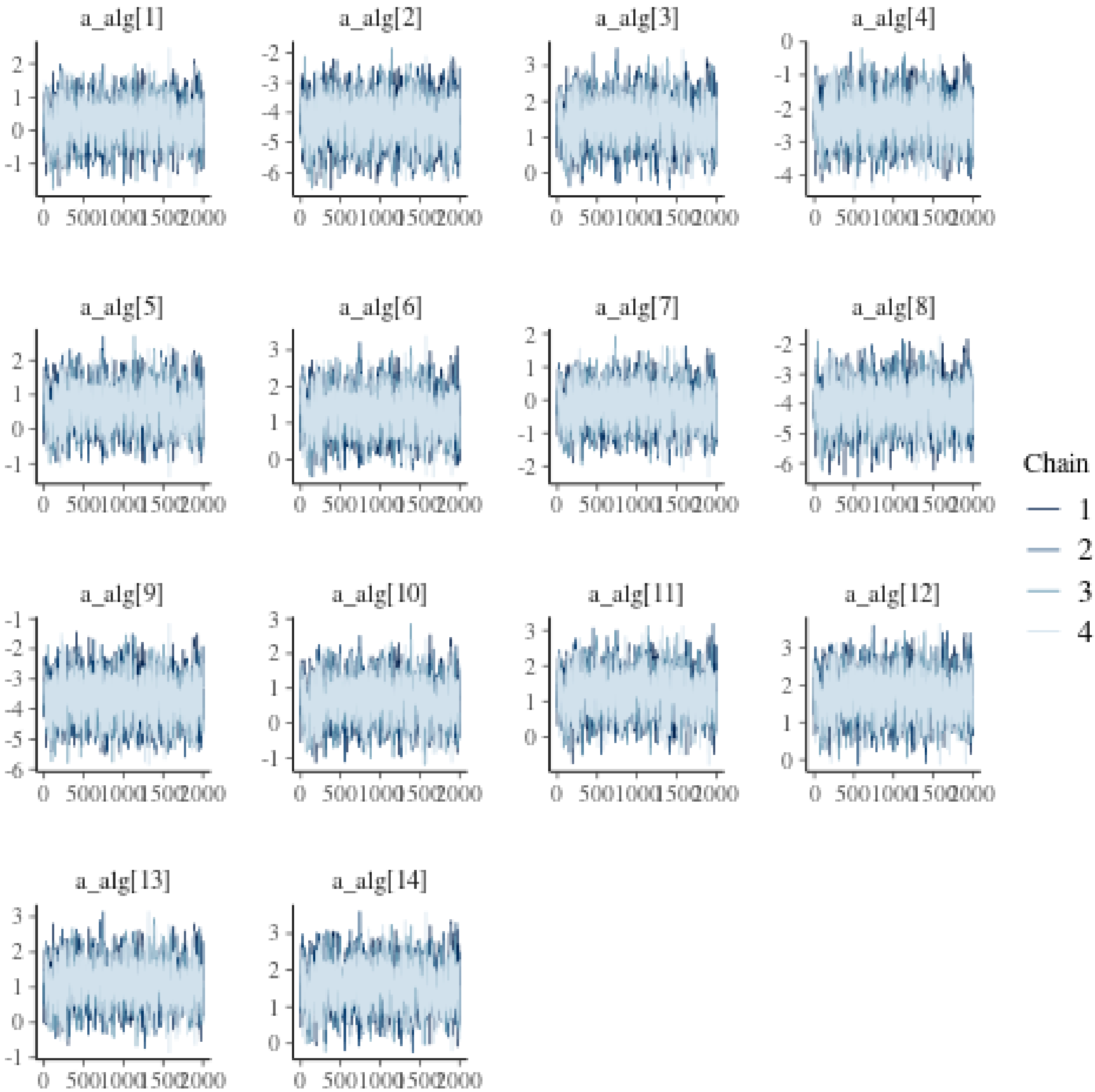

FIGURE 3. Simulated MCMC chains for the strength parameters a_i (25%)

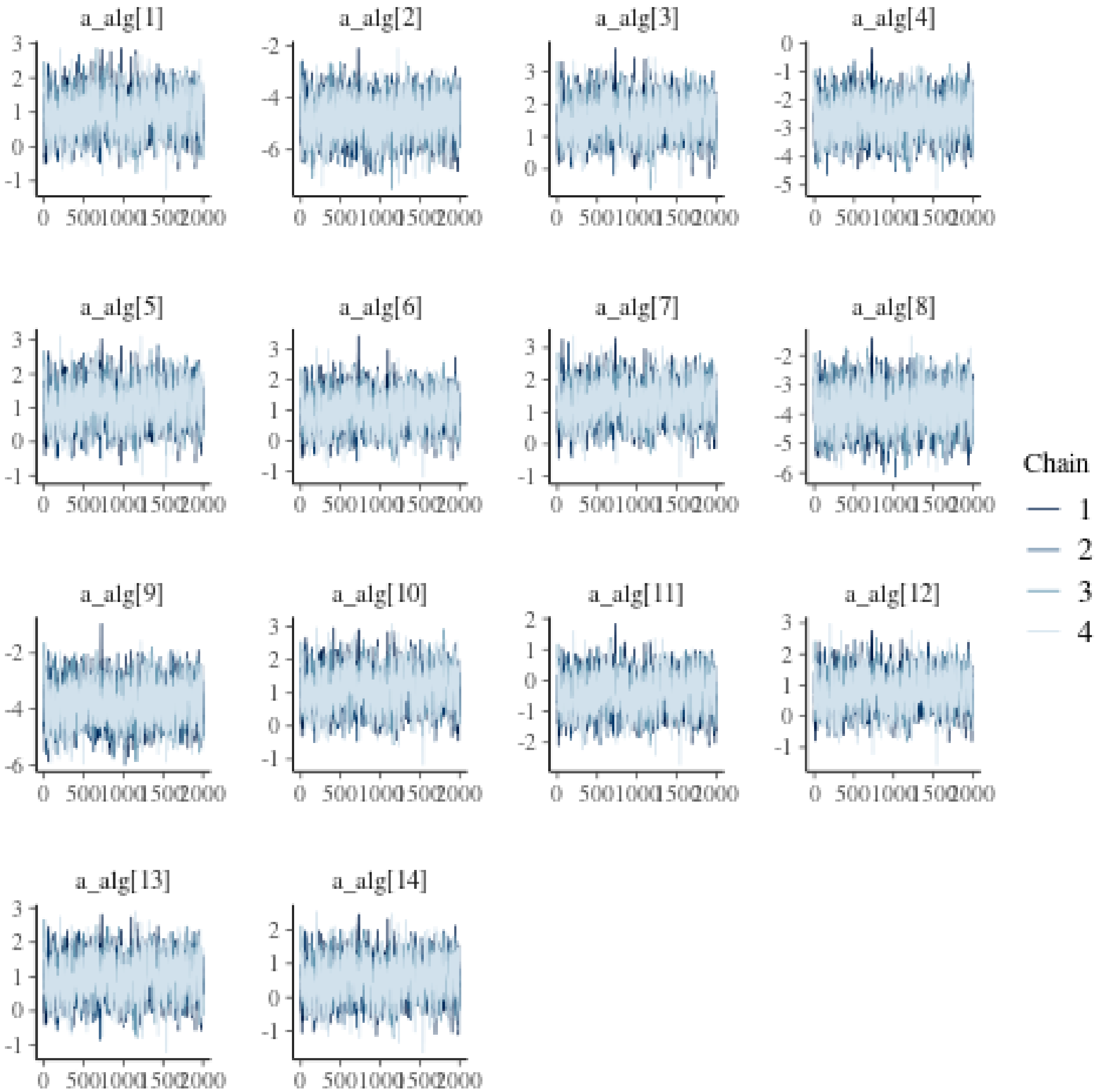

FIGURE 4. Simulated MCMC chains for the strength parameters a_i (50%)

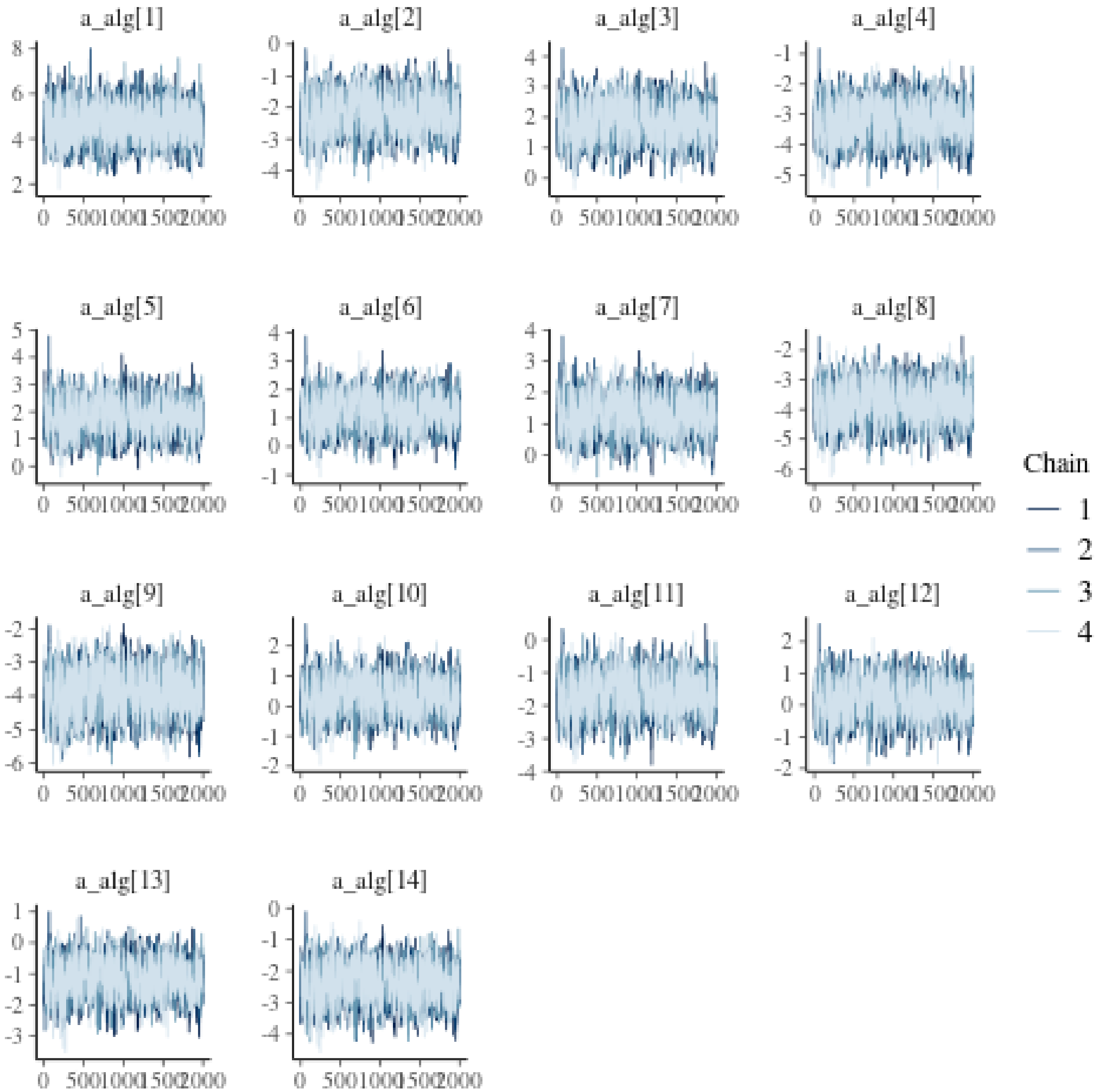

FIGURE 5. Simulated MCMC chains for the strength parameters a_i (75%)

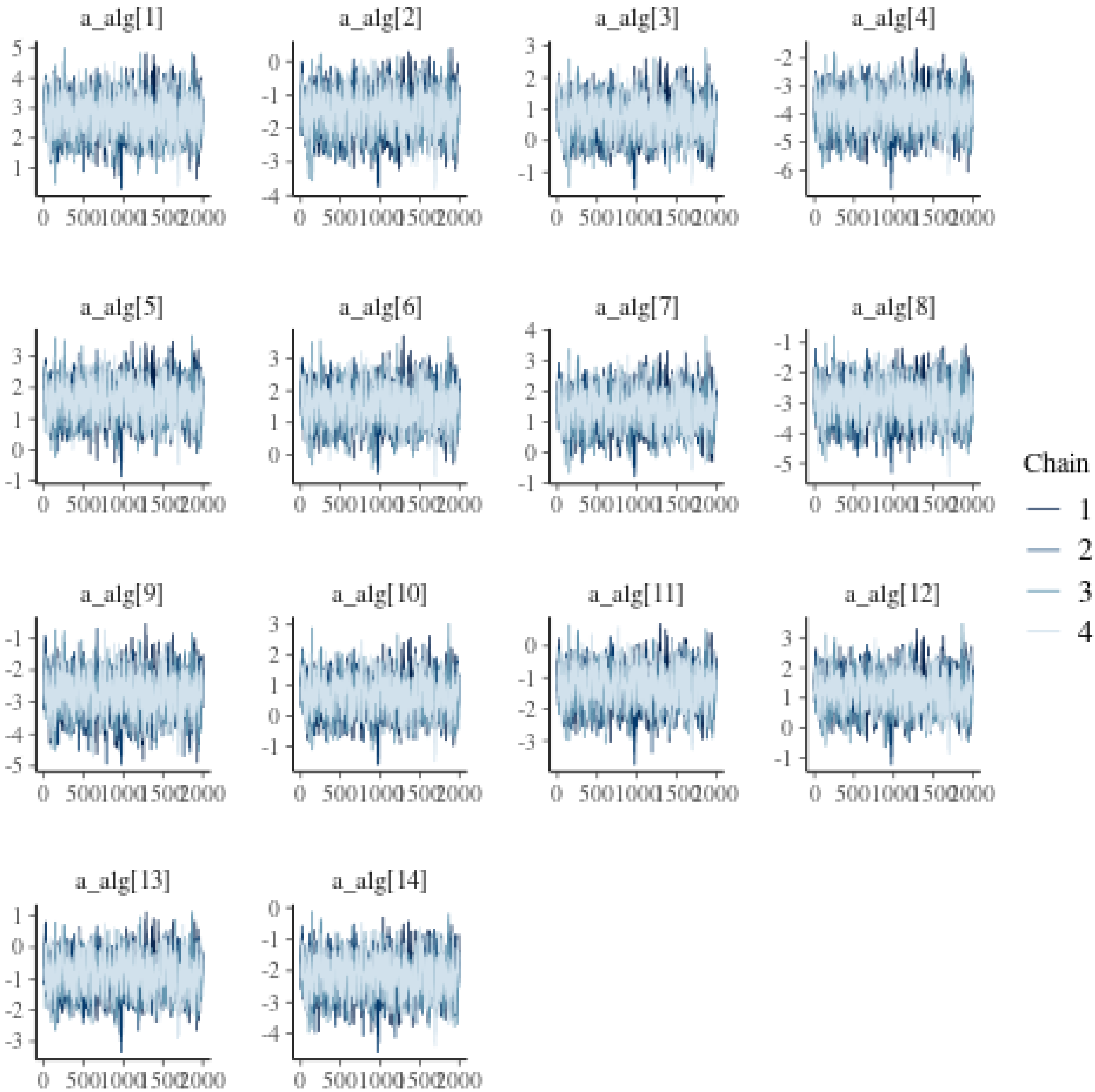

FIGURE 6. Simulated MCMC chains for the strength parameters a_i (90%)

Table 2. Diagnostics for 10% of available labels

Variable	Rhat	ess_bulk	ess_tail
a_alg[1]	1.006	930.781	1620.641
a_alg[2]	1.003	1420.982	2790.689
a_alg[3]	1.006	849.576	1400.238
a_alg[4]	1.005	1076.228	2028.801
a_alg[5]	1.006	837.920	1447.750
a_alg[6]	1.006	862.660	1539.259
a_alg[7]	1.006	854.106	1469.088
a_alg[8]	1.003	1247.605	2262.456
a_alg[9]	1.004	1249.388	2406.563
a_alg[10]	1.006	870.370	1455.832
a_alg[11]	1.005	946.641	1693.693
a_alg[12]	1.006	878.207	1633.142
a_alg[13]	1.006	853.407	1470.405
a_alg[14]	1.006	860.742	1406.467

Table 3. Diagnostics for 25% of available labels

Variable	Rhat	ess_bulk	ess_tail
a_alg[1]	1.006	875.652	1328.454
a_alg[2]	1.003	1184.027	2415.331
a_alg[3]	1.006	869.503	1297.981
a_alg[4]	1.004	1000.064	1875.254
a_alg[5]	1.005	880.397	1360.262
a_alg[6]	1.006	845.098	1295.582
a_alg[7]	1.006	863.251	1171.705
a_alg[8]	1.003	1091.075	2146.680
a_alg[9]	1.004	1102.401	2122.005
a_alg[10]	1.007	860.174	1260.263
a_alg[11]	1.006	874.051	1320.953
a_alg[12]	1.006	839.967	1217.099
a_alg[13]	1.007	844.654	1170.443
a_alg[14]	1.006	856.543	1352.789

Table 4. Diagnostics for 50% of available labels

Variable	Rhat	ess_bulk	ess_tail
a_alg[1]	1.006	923.599	1536.612
a_alg[2]	1.002	1541.398	3235.483
a_alg[3]	1.006	953.080	1648.550
a_alg[4]	1.003	1229.076	2334.682
a_alg[5]	1.007	917.699	1611.186
a_alg[6]	1.006	949.147	1707.167
a_alg[7]	1.006	922.449	1679.524
a_alg[8]	1.003	1373.961	2533.233
a_alg[9]	1.003	1349.500	2575.455
a_alg[10]	1.006	935.044	1529.858
a_alg[11]	1.005	988.905	1715.657
a_alg[12]	1.006	937.424	1673.032
a_alg[13]	1.006	946.234	1541.462
a_alg[14]	1.006	938.252	1564.519

Table 5. Diagnostic for 75% of available labels

Variable	Rhat	ess_bulk	ess_tail
a_alg[1]	1.001	1605.863	3034.650
a_alg[2]	1.002	844.253	1205.099
a_alg[3]	1.002	911.582	1497.275
a_alg[4]	1.002	893.164	1342.419
a_alg[5]	1.003	885.522	1560.613
a_alg[6]	1.002	840.815	1442.053
a_alg[7]	1.003	846.997	1401.868
a_alg[8]	1.002	942.949	1438.113
a_alg[9]	1.001	955.214	1473.940
a_alg[10]	1.002	821.244	1319.835
a_alg[11]	1.002	825.454	1179.804
a_alg[12]	1.003	804.589	1148.154
a_alg[13]	1.002	833.639	1182.323
a_alg[14]	1.002	863.293	1155.661

Table 6. Diagnostics for 90% of available labels

Variable	Rhat	ess_bulk	ess_tail
a_alg[1]	1.004	636.840	1216.352
a_alg[2]	1.004	601.977	1152.796
a_alg[3]	1.004	586.138	1145.326
a_alg[4]	1.003	710.298	1281.454
a_alg[5]	1.004	581.250	1178.959
a_alg[6]	1.004	590.042	1131.033
a_alg[7]	1.004	598.670	1068.376
a_alg[8]	1.004	648.385	1322.970
a_alg[9]	1.003	643.530	1032.394
a_alg[10]	1.004	586.434	1089.956
a_alg[11]	1.004	616.109	1260.695
a_alg[12]	1.004	583.490	1206.982
a_alg[13]	1.004	590.859	1230.768
a_alg[14]	1.004	604.542	1218.919

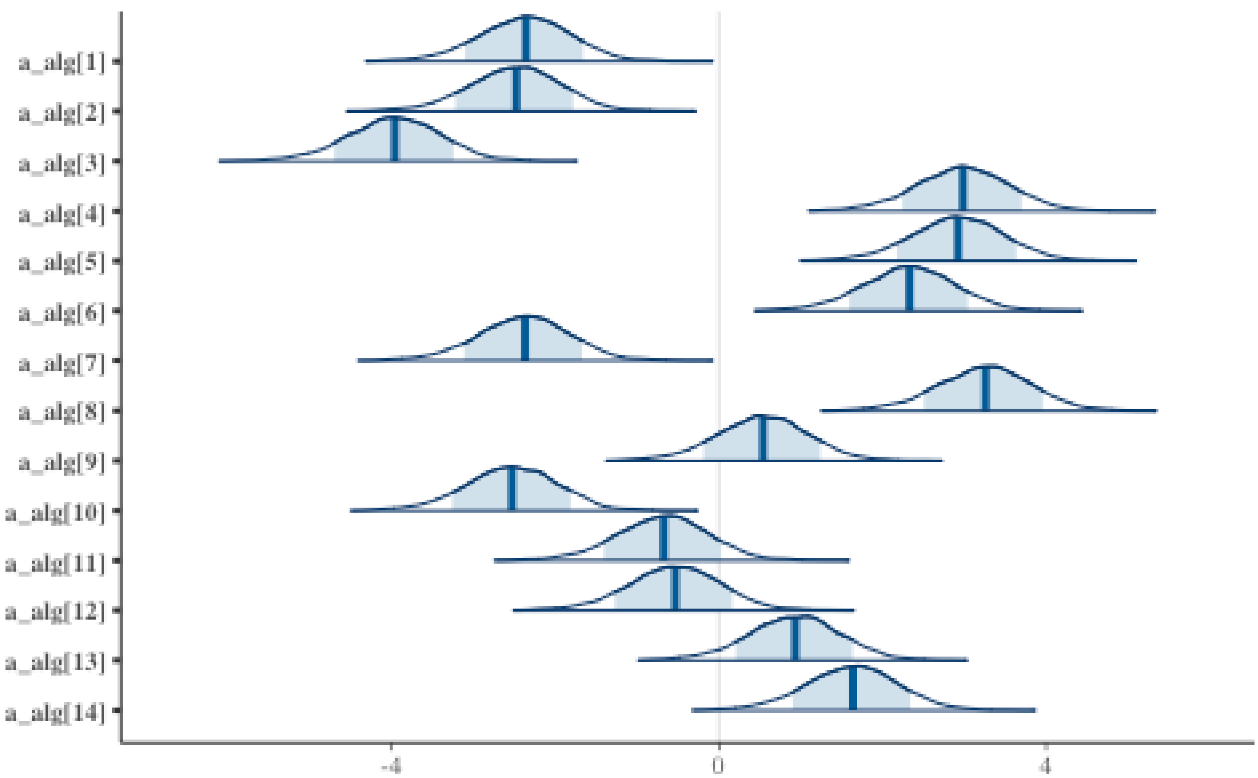

Figure 7. The estimated posterior distributions of the strength parameters a_i. The dark-blue line is the mean value of a_i and the light-blue shaded area is the HPD Interval.

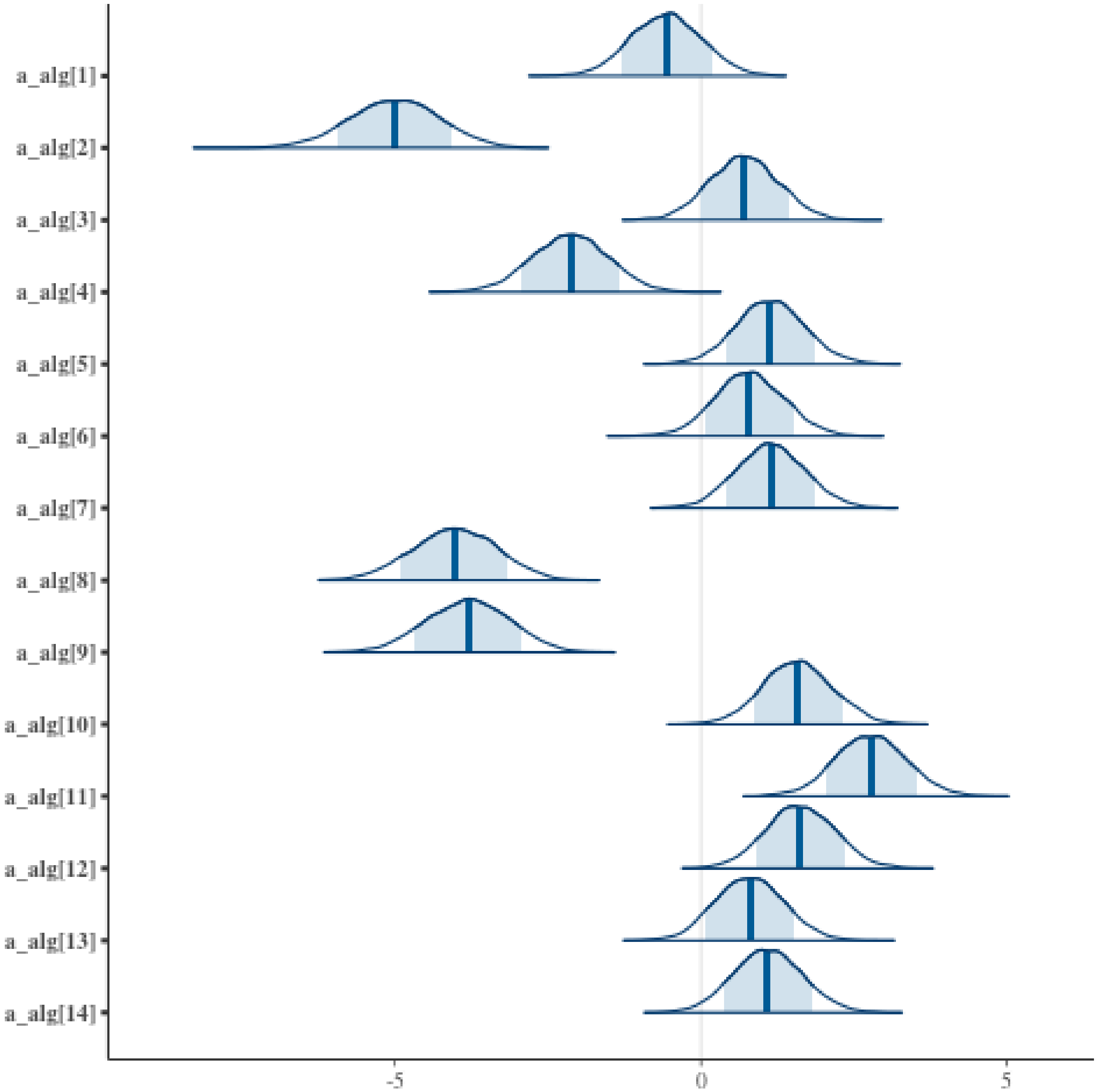

FIGURE 8. The estimated posterior distribution of the strength parameters a_i for 10% of labeled data. The dark-blue line is the mean value of a_i and the light-blue shaded area is the HPD Interval.

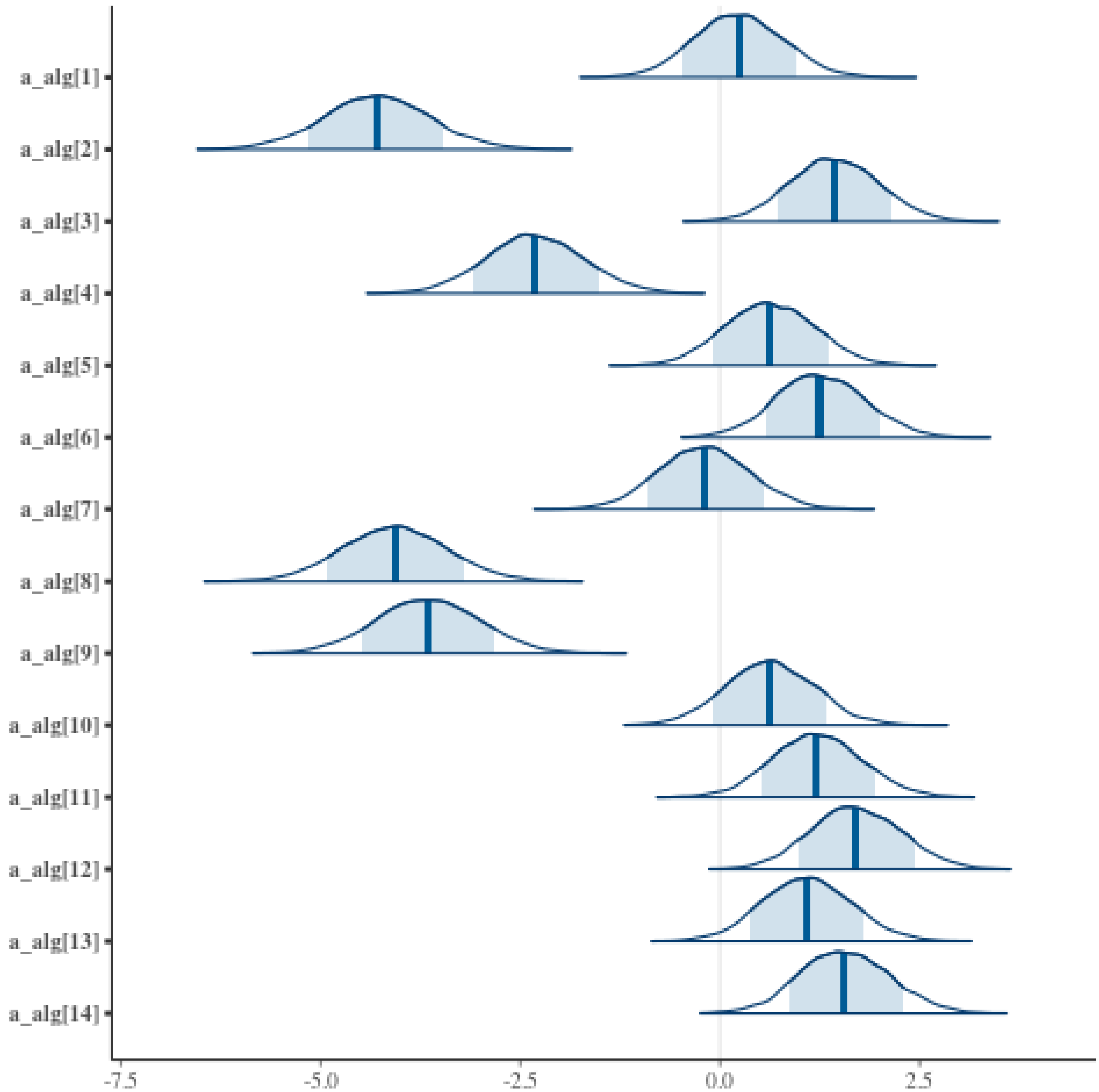

FIGURE 9. The estimated posterior distribution of the strength parameters a_i for 25% of labeled data. The dark-blue line is the mean value of a_i and the light-blue shaded area is the HPD Interval.

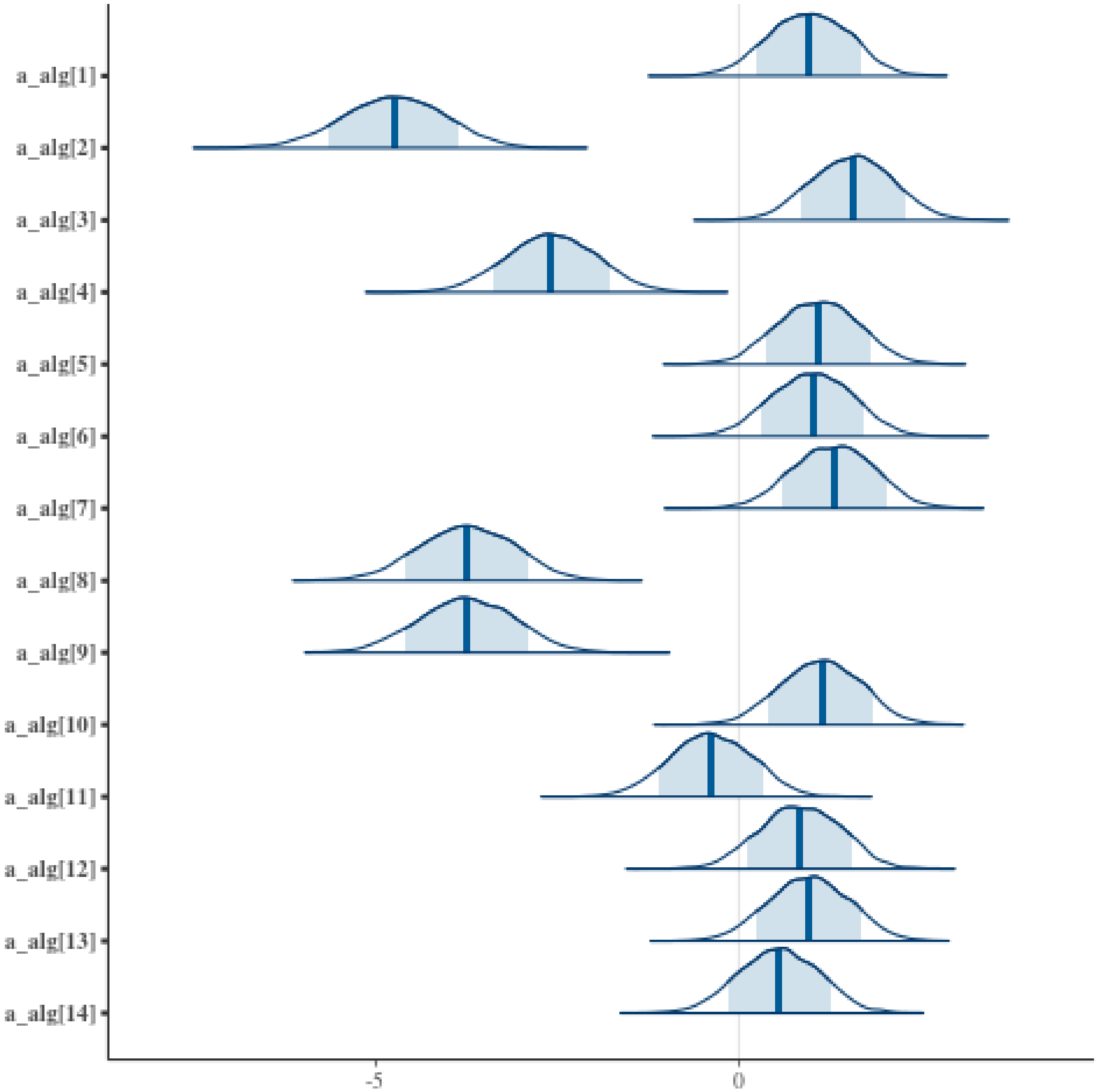

FIGURE 10. The estimated posterior distribution of the strength parameters a_i for 50% of labeled data. The dark-blue line is the mean value of a_i and the light-blue shaded area is the HPD Interval.

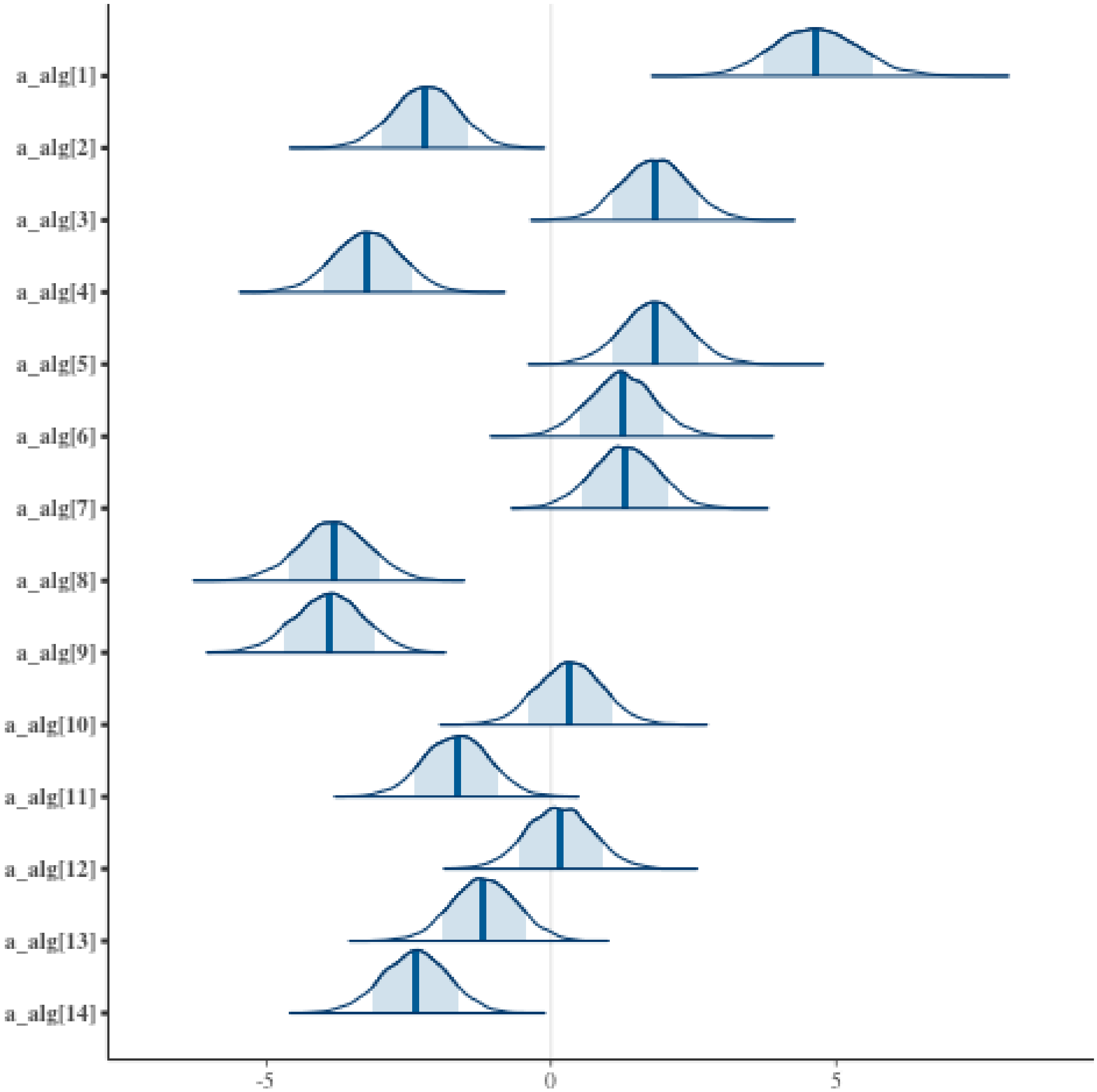

FIGURE 11. The estimated posterior distribution of the strength parameters a_i for 75% of labeled data. The dark-blue line is the mean value of a_i and the light-blue shaded area is the HPD Interval.

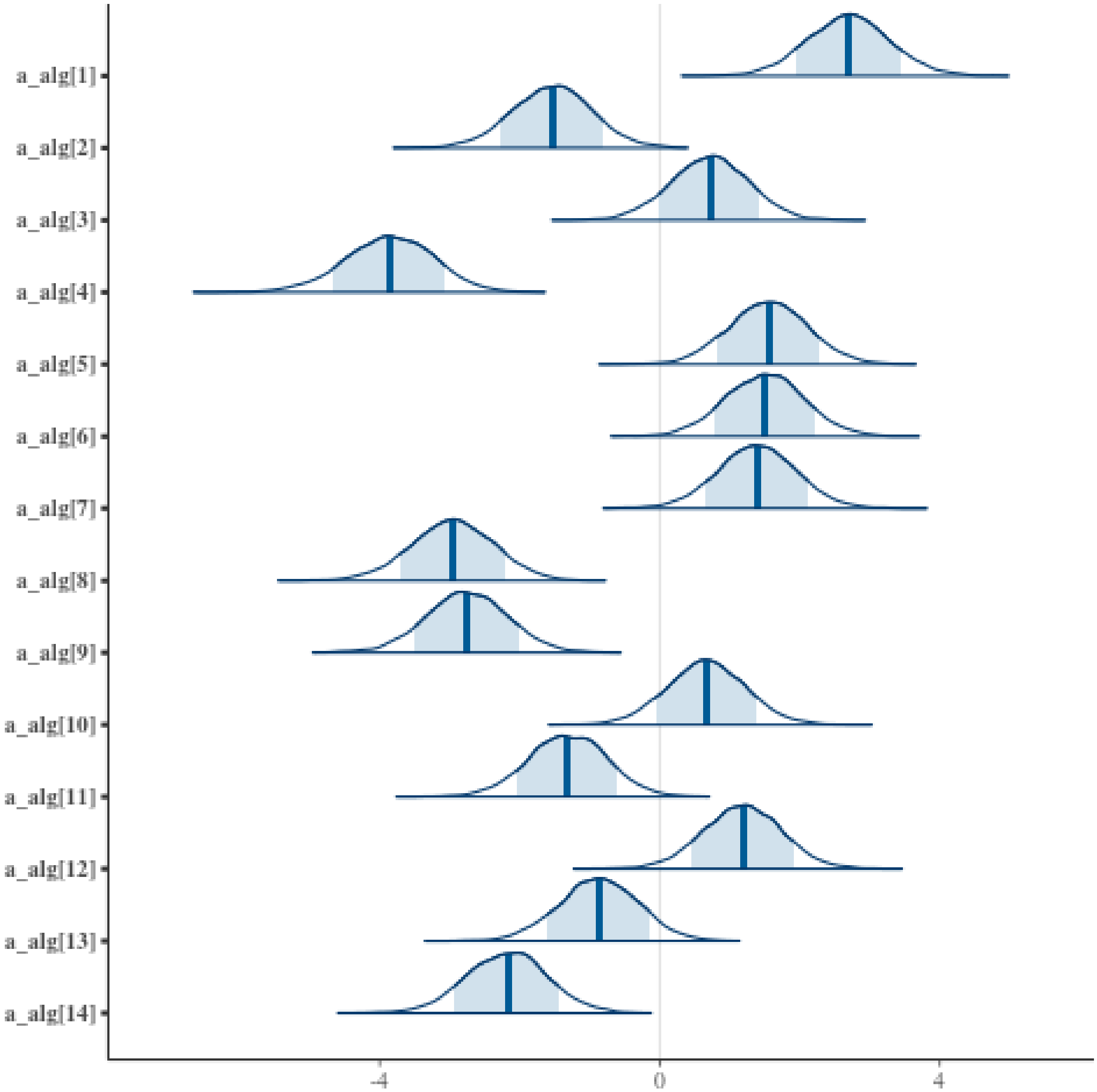

FIGURE 12. The estimated posterior distribution of the strength parameters a_i for 90% of labeled data. The dark-blue line is the mean value of a_i and the light-blue shaded area is the HPD Interval.

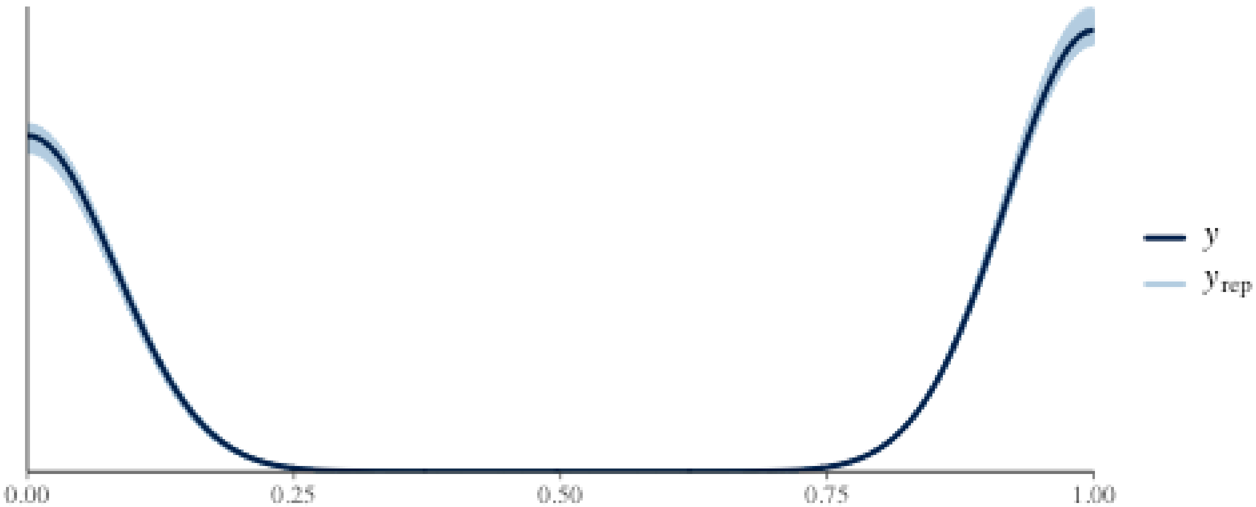

FIGURE 13. Kernel density estimate of the observed dataset y with density estimates for the simulated datasets y_{rep}.

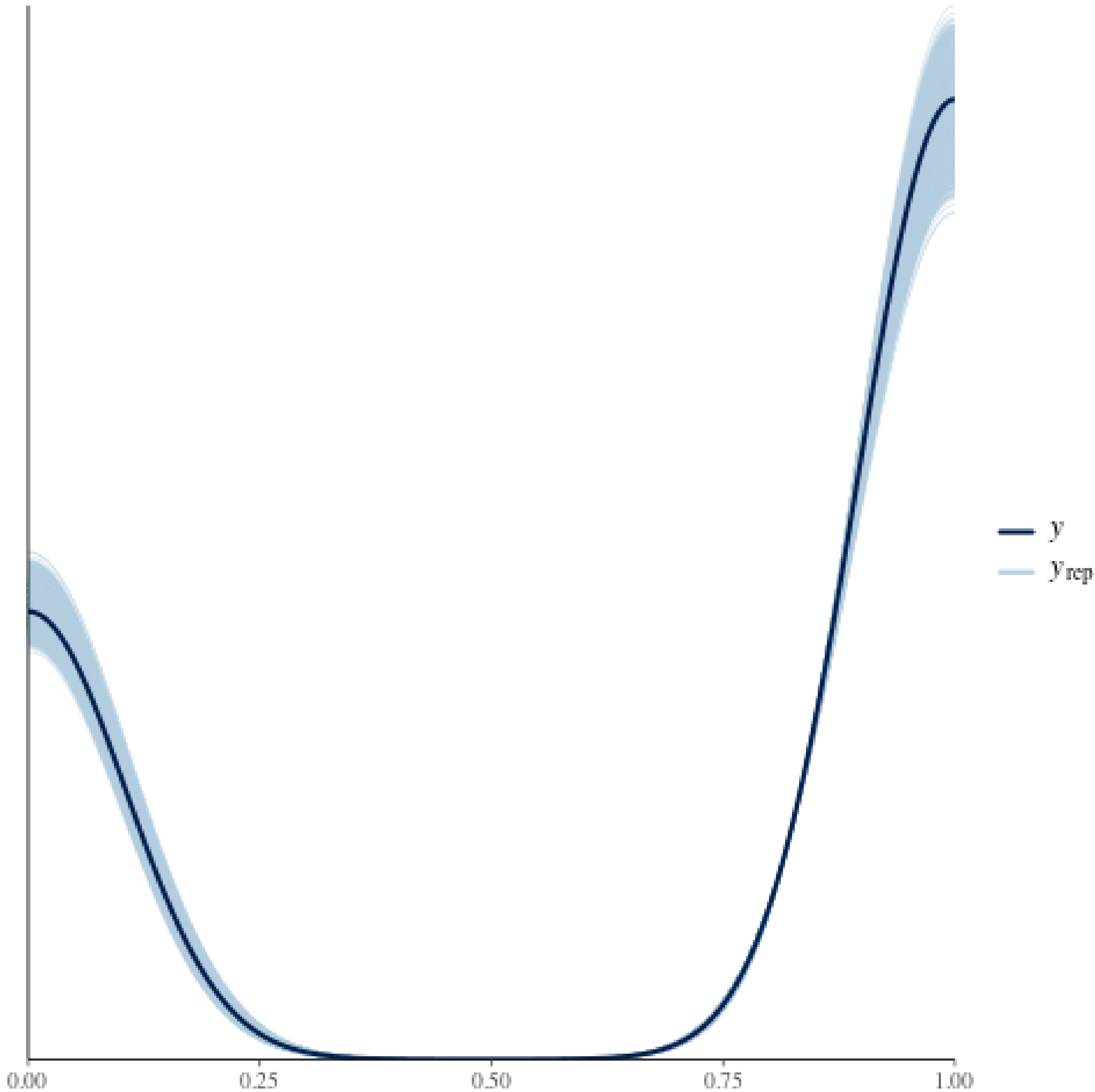

Figure 14. Kernel density estimate of the observed dataset y with density estimates for the simulated datasets y_{rep} when having access to 10% of labels

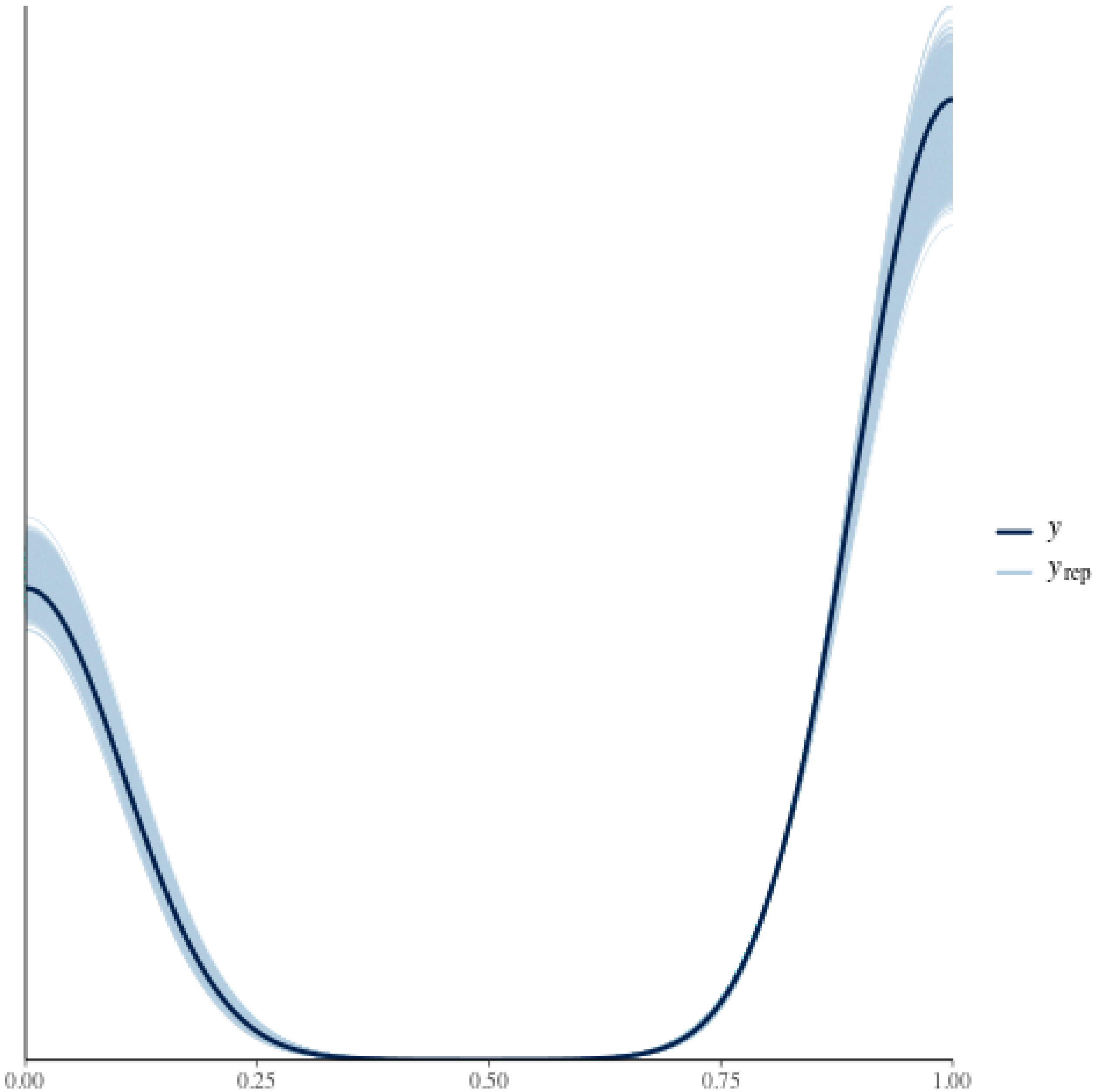

FIGURE 15. Kernel density estimate of the observed dataset y with density estimates for the simulated datasets y_{rep} when having access to 25% of labels

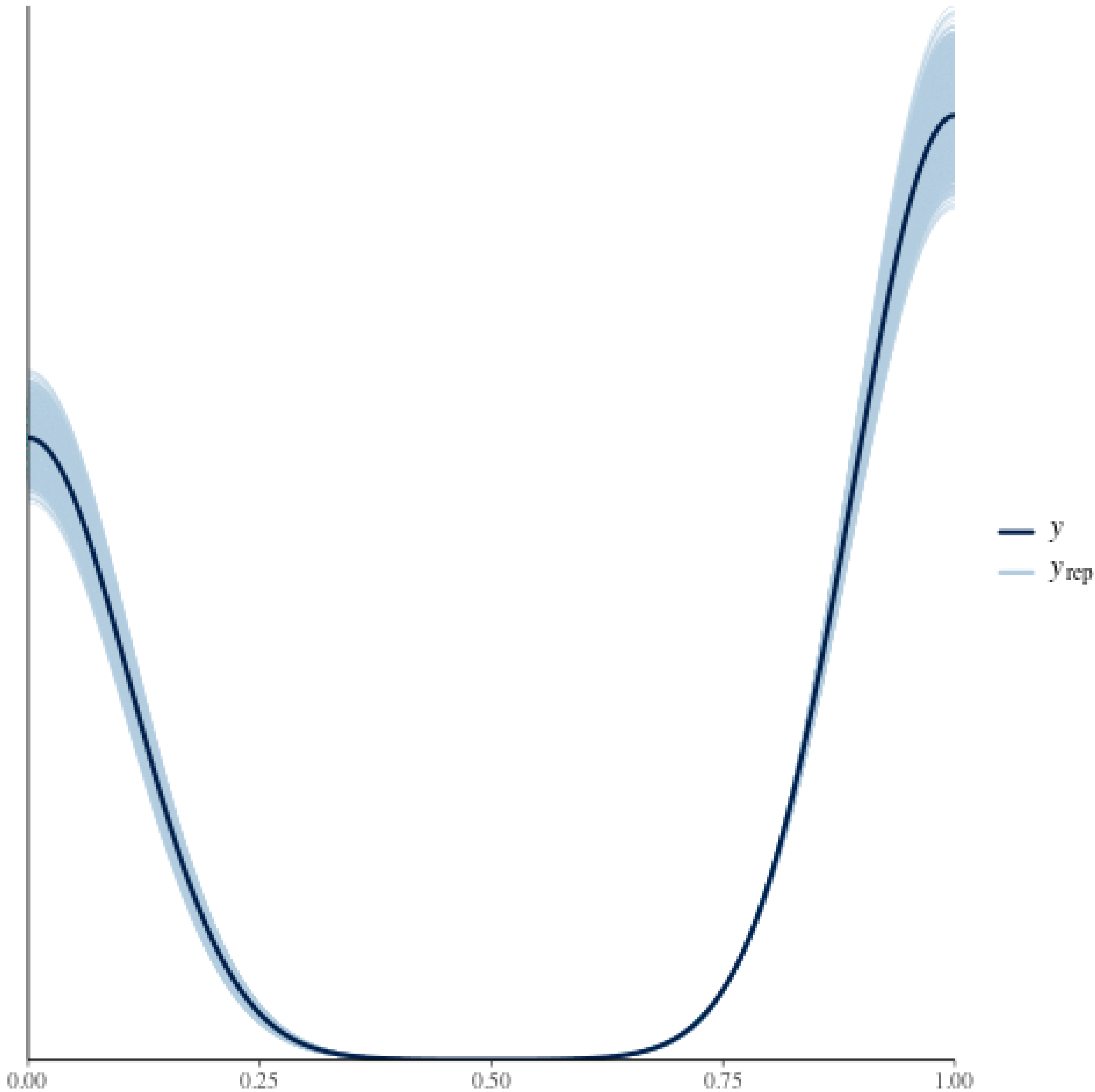

Figure 16. Kernel density estimate of the observed dataset y with density estimates for the simulated datasets y_{rep} when having access to 25% of labels

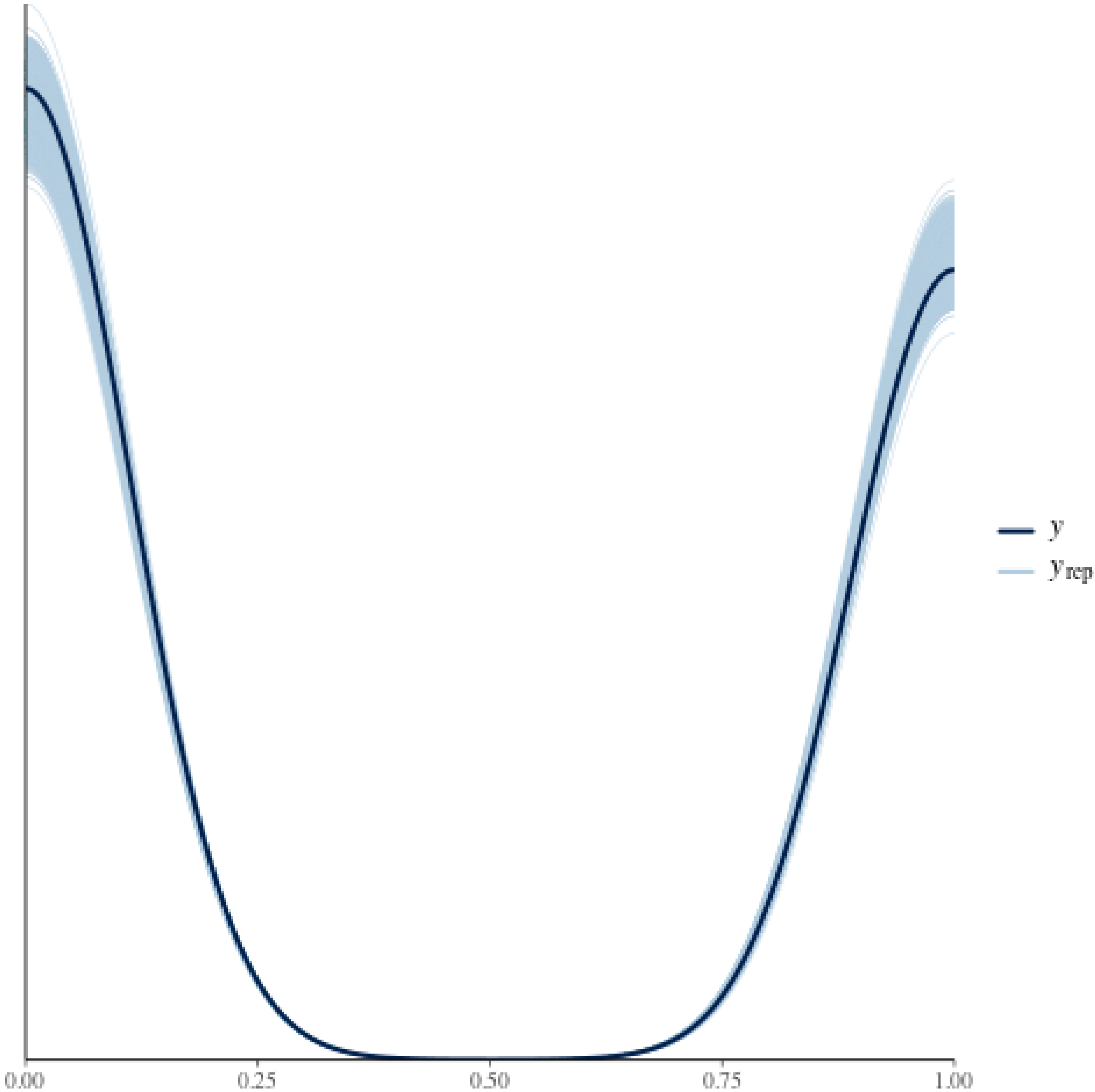

FIGURE 17. Kernel density estimate of the observed dataset y with density estimates for the simulated datasets y_{rep} when having access to 75% of labels

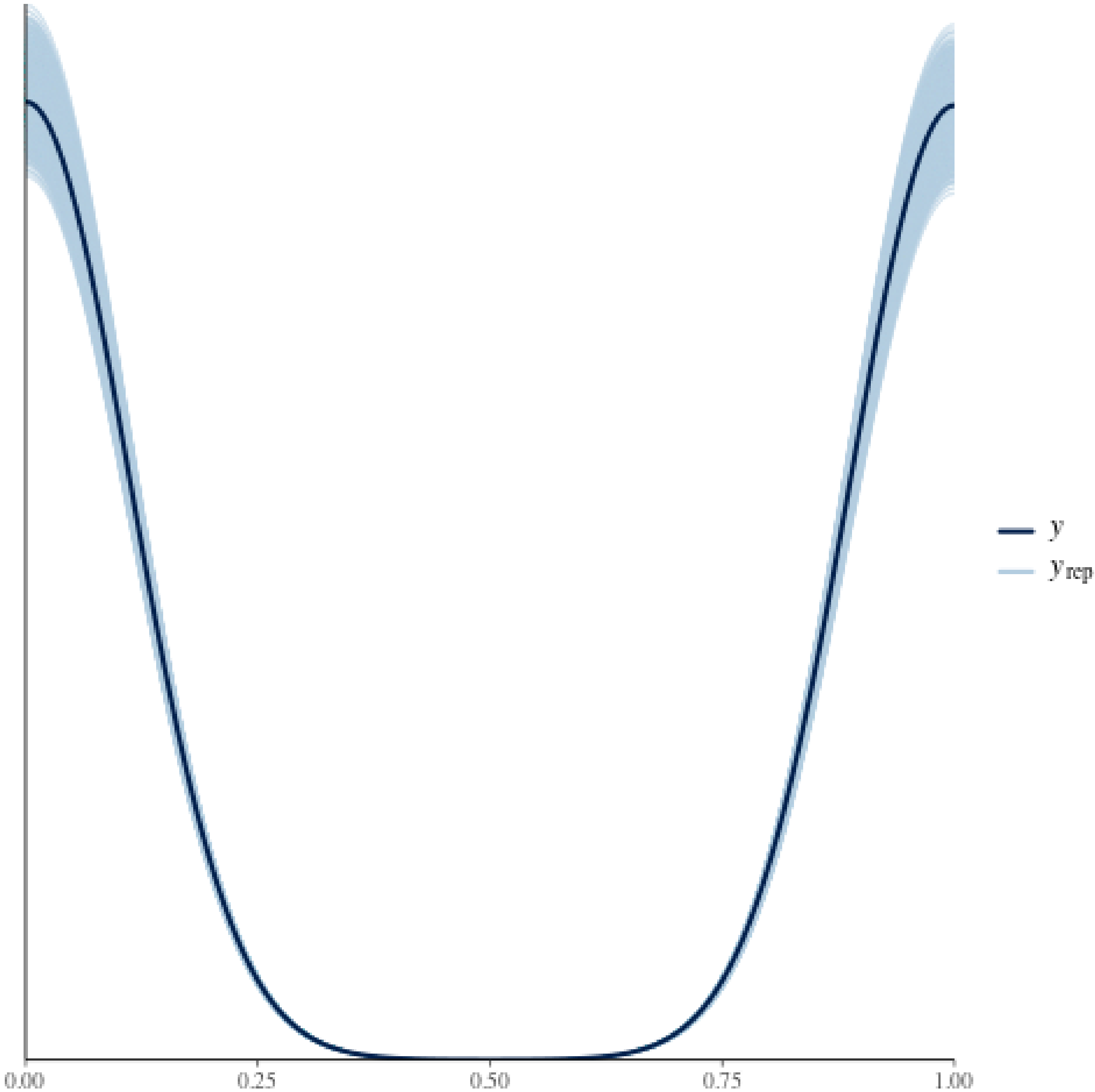

FIGURE 18. Kernel density estimate of the observed dataset y with density estimates for the simulated datasets y_{rep} when having access to 90% of labels

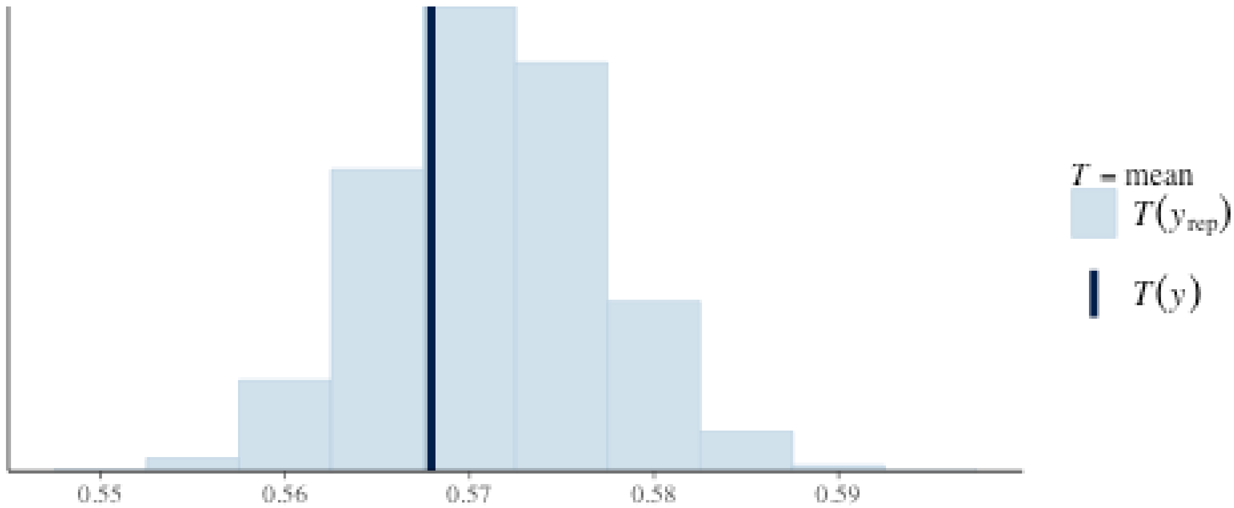

FIGURE 19. Histograms the mean(y_{rep}) computed from the posterior predictive distribution. The dark-blue line is computed from the observed data.

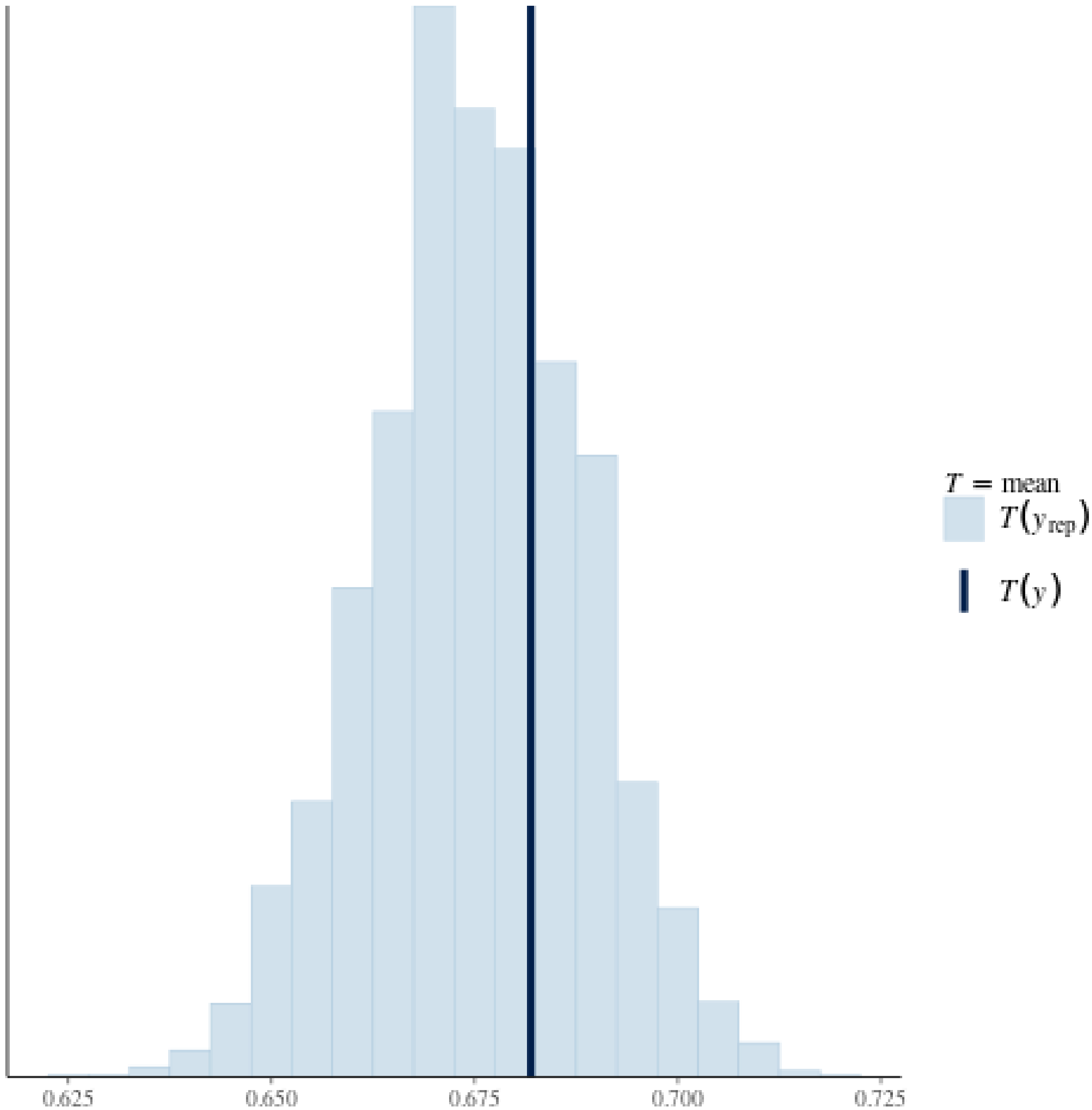

FIGURE 20. Histograms the mean(y_{rep}) computed from the posterior predictive distribution. The dark-blue line is computed from the observed data. (10% of available labels)

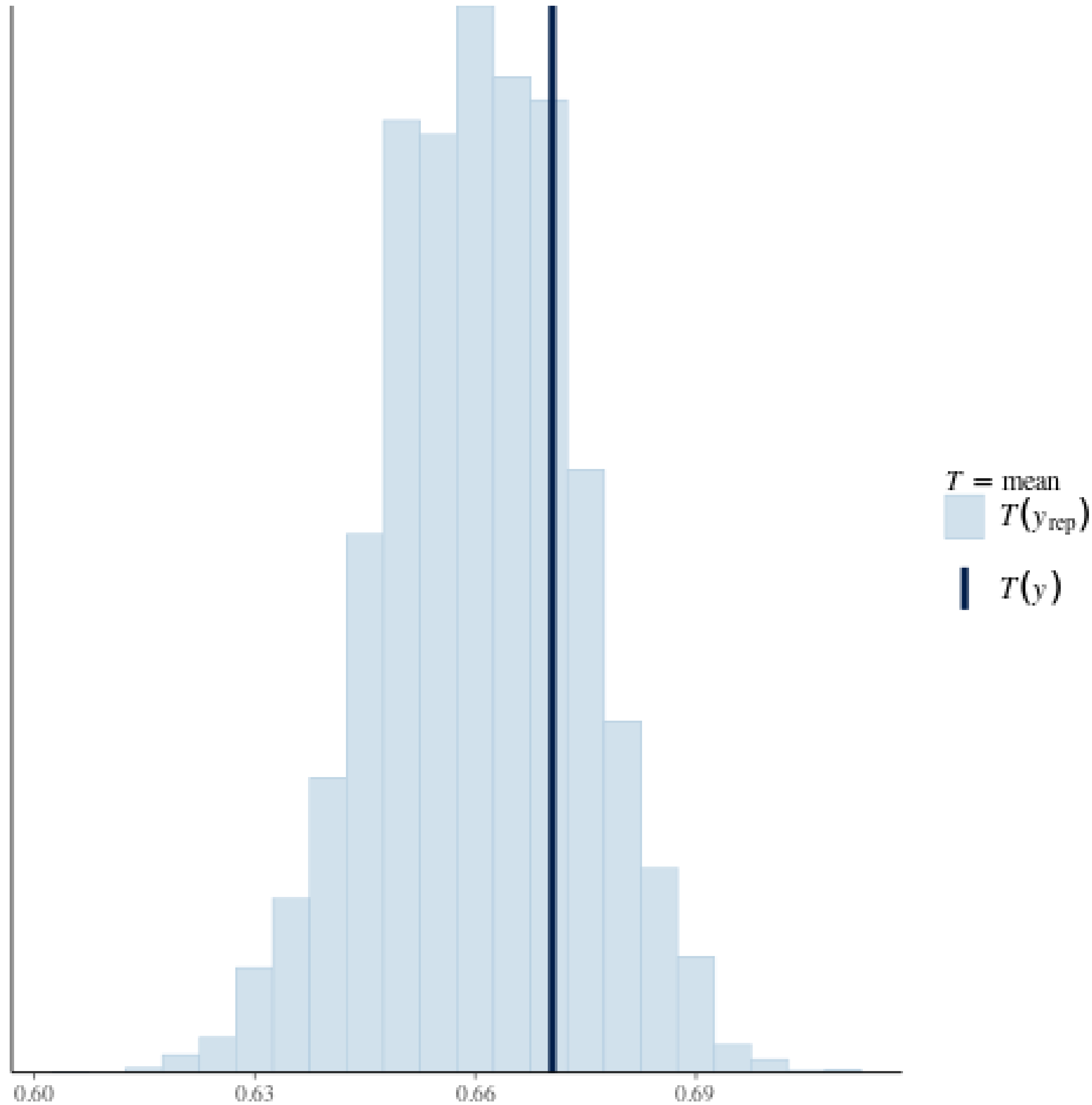

FIGURE 21. Histograms the mean(y_{rep}) computed from the posterior predictive distribution. The dark-blue line is computed from the observed data. (25% of available labels)

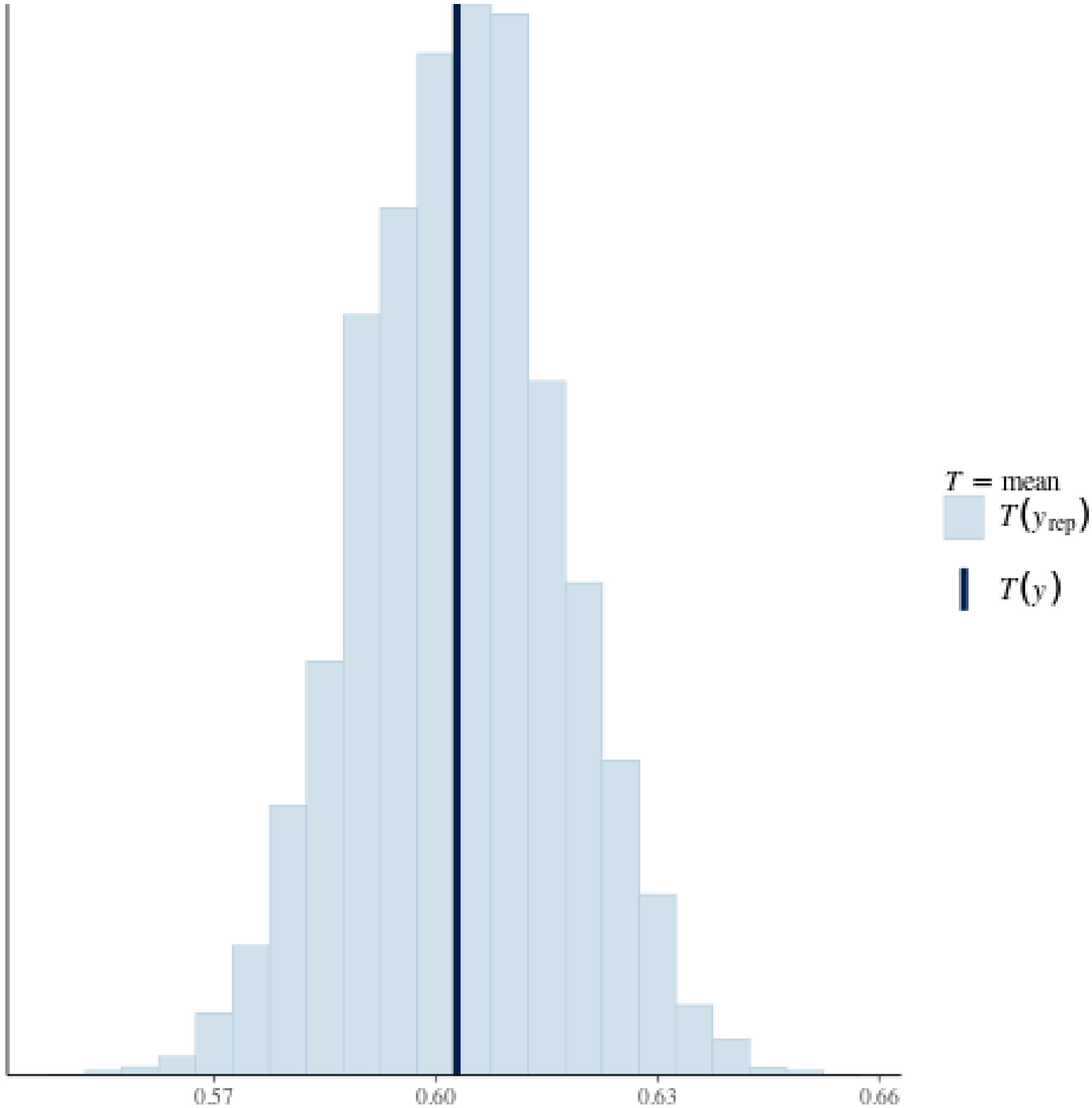

Figure 22. Histograms the mean(y_{rep}) computed from the posterior predictive distribution. The dark-blue line is computed from the observed data. (50% of available labels)

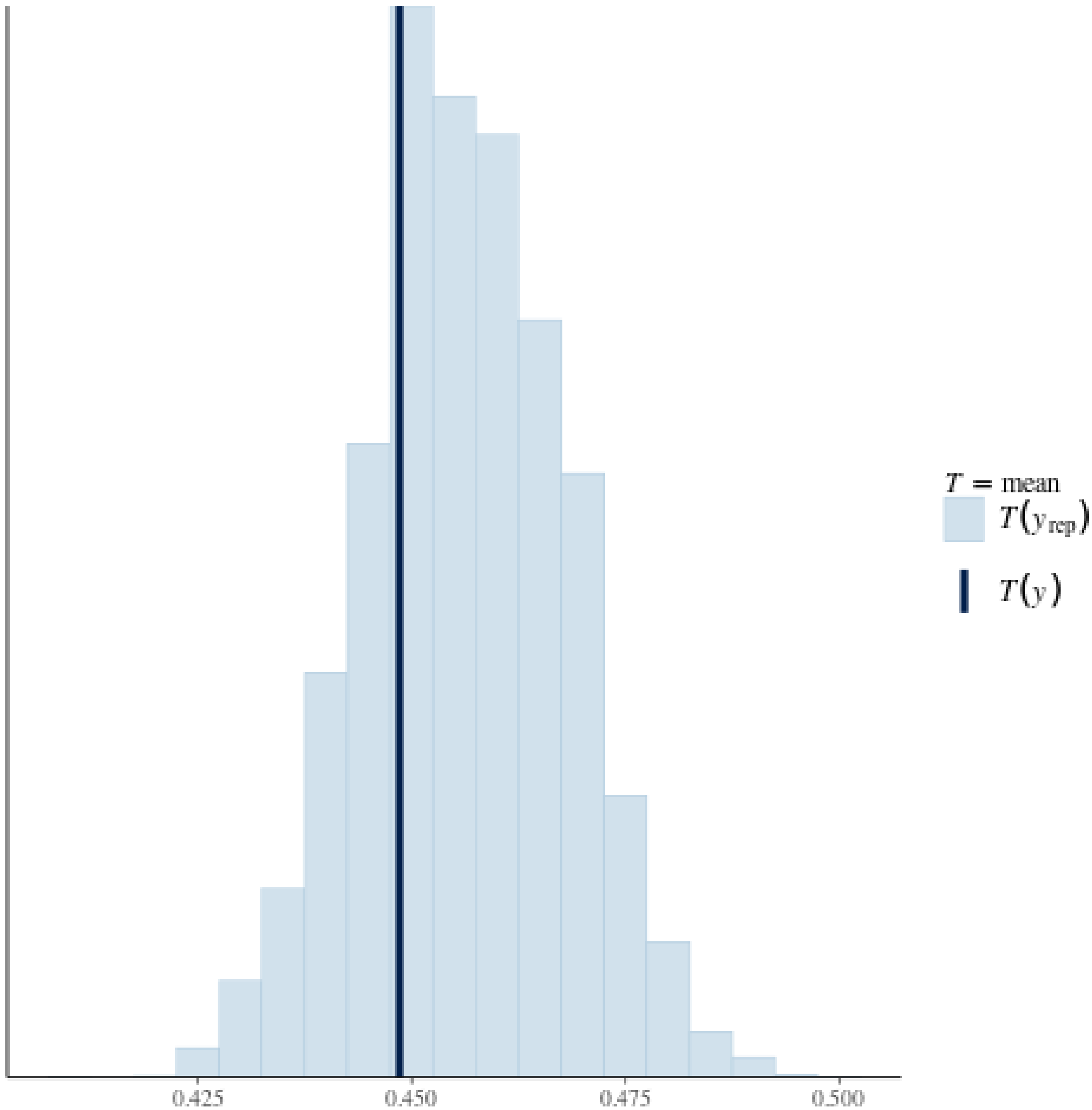

FIGURE 23. Histograms the mean(y_{rep}) computed from the posterior predictive distribution. The dark-blue line is computed from the observed data. 75% of available labels)

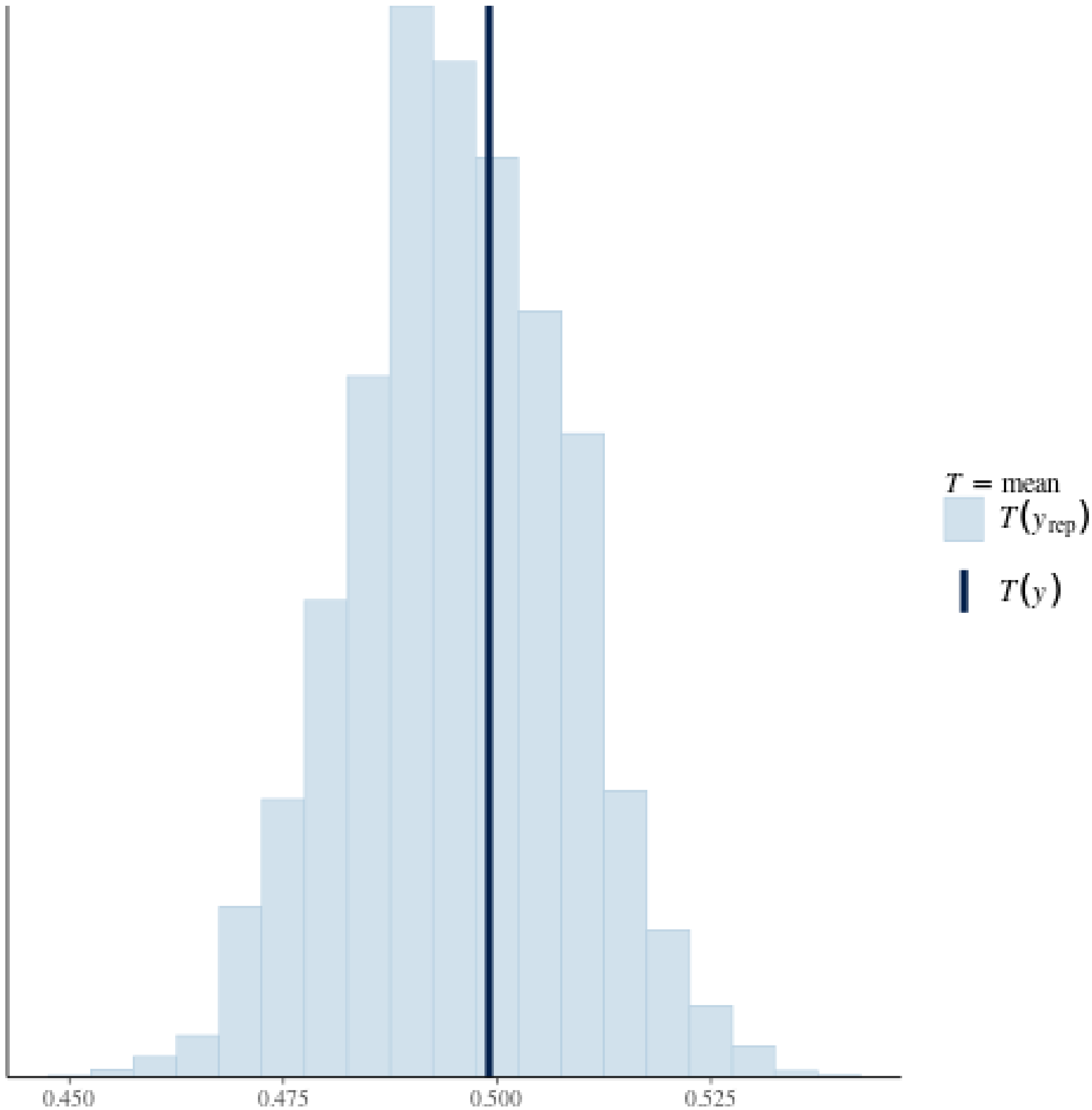

FIGURE 24. Histograms the mean(y_{rep}) computed from the posterior predictive distribution. The dark-blue line is computed from the observed data. (90% of available labels)

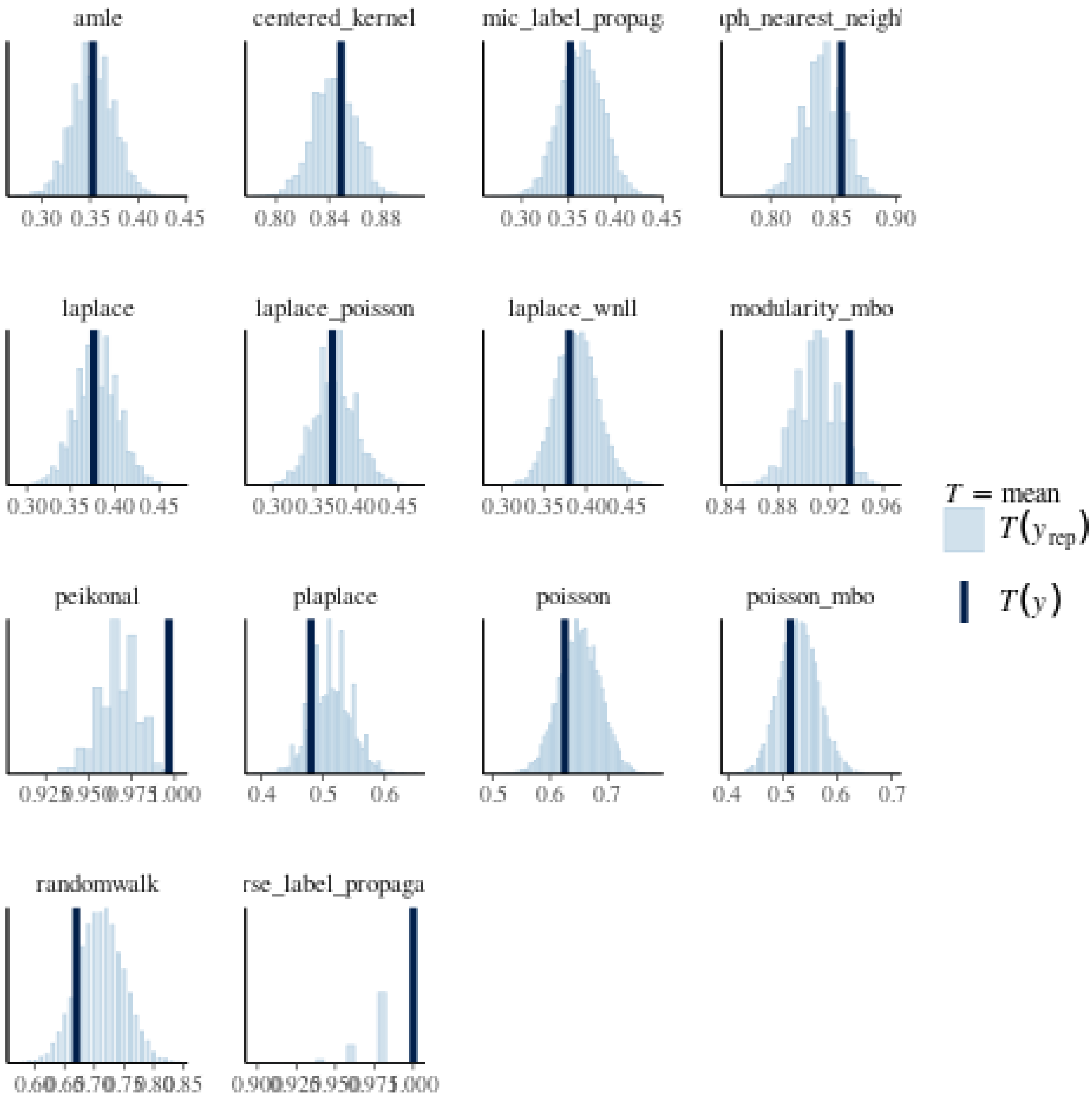

FIGURE 25. Checking posterior predictive (aggregated) test statistics within the region, in this case, the mean. The vertical lines are observed means.

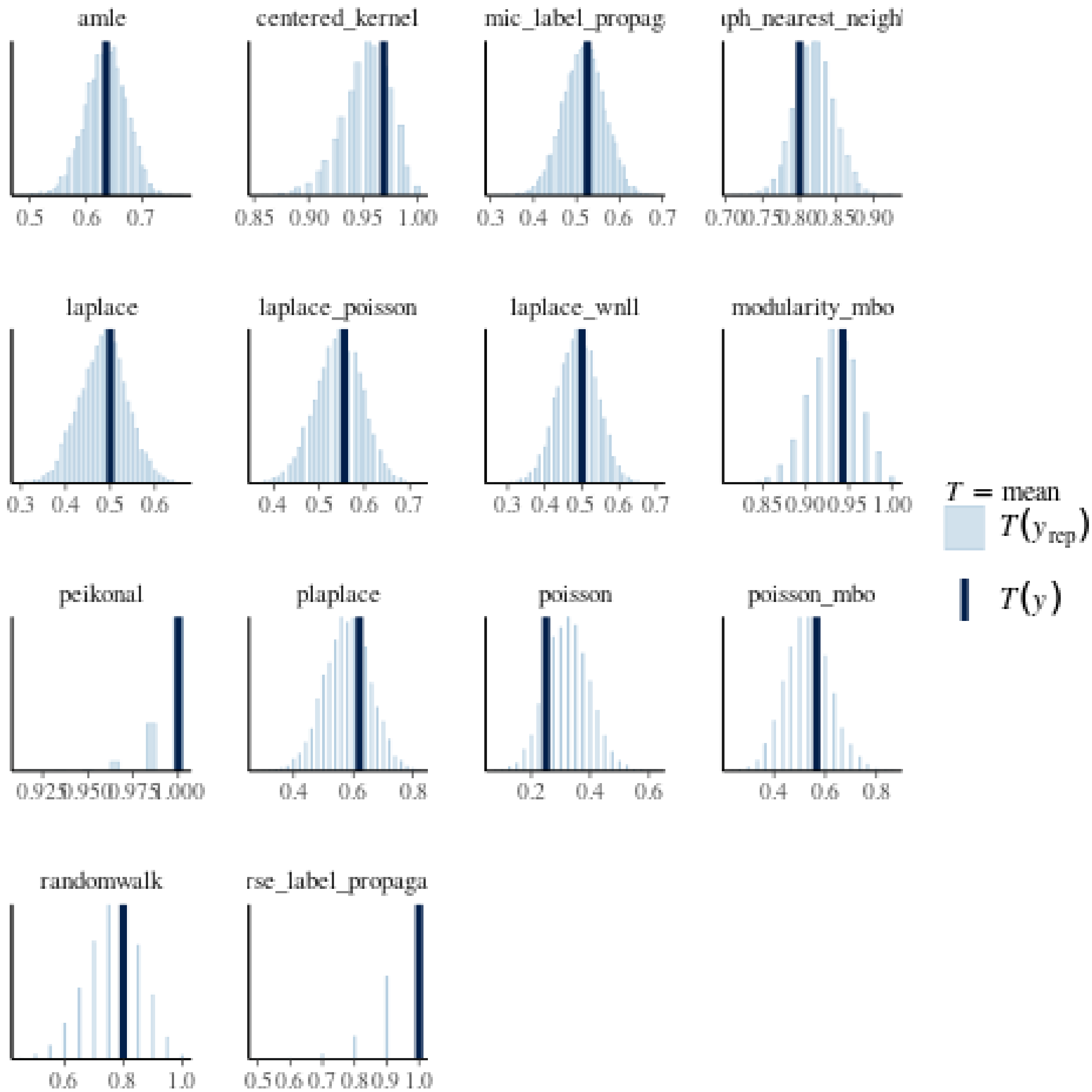

FIGURE 26. Checking posterior predictive (10%) test statistics within the region, in this case, the mean. The vertical lines are observed means.

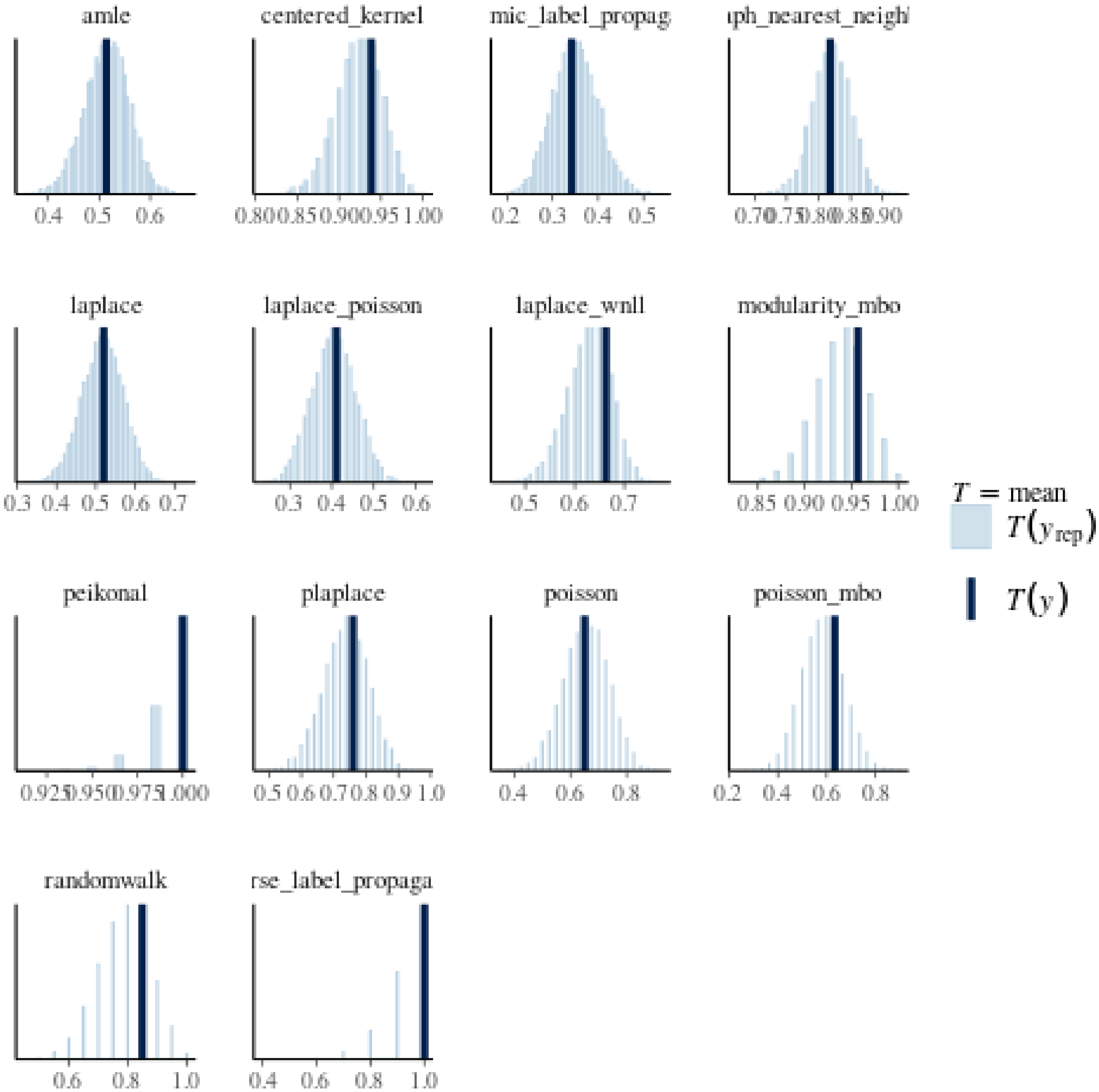

FIGURE 27. Checking posterior predictive (25%) test statistics within the region, in this case, the mean. The vertical lines are observed means.

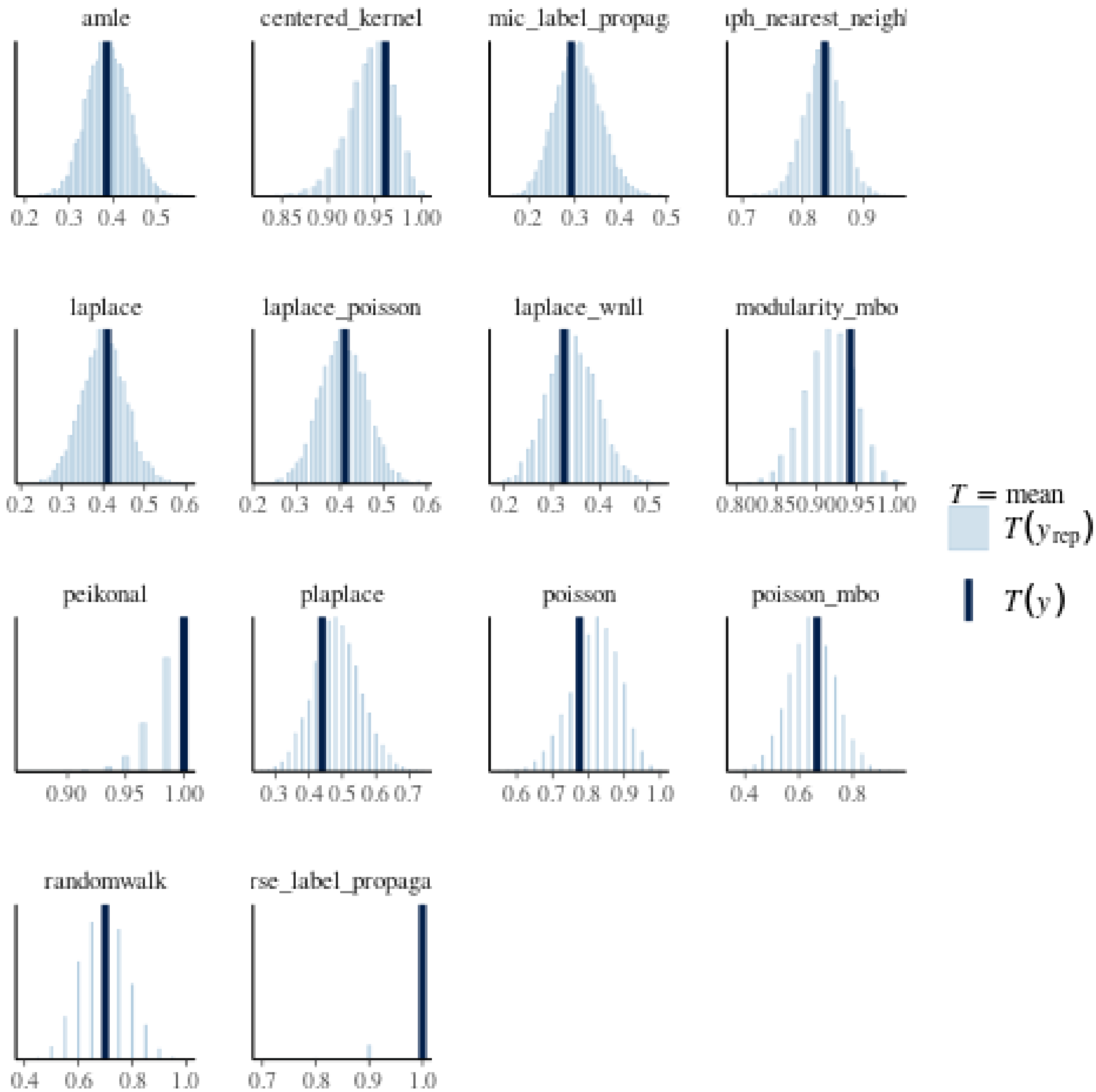

FIGURE 28. Checking posterior predictive (50%) test statistics within the region, in this case, the mean. The vertical lines are observed means.

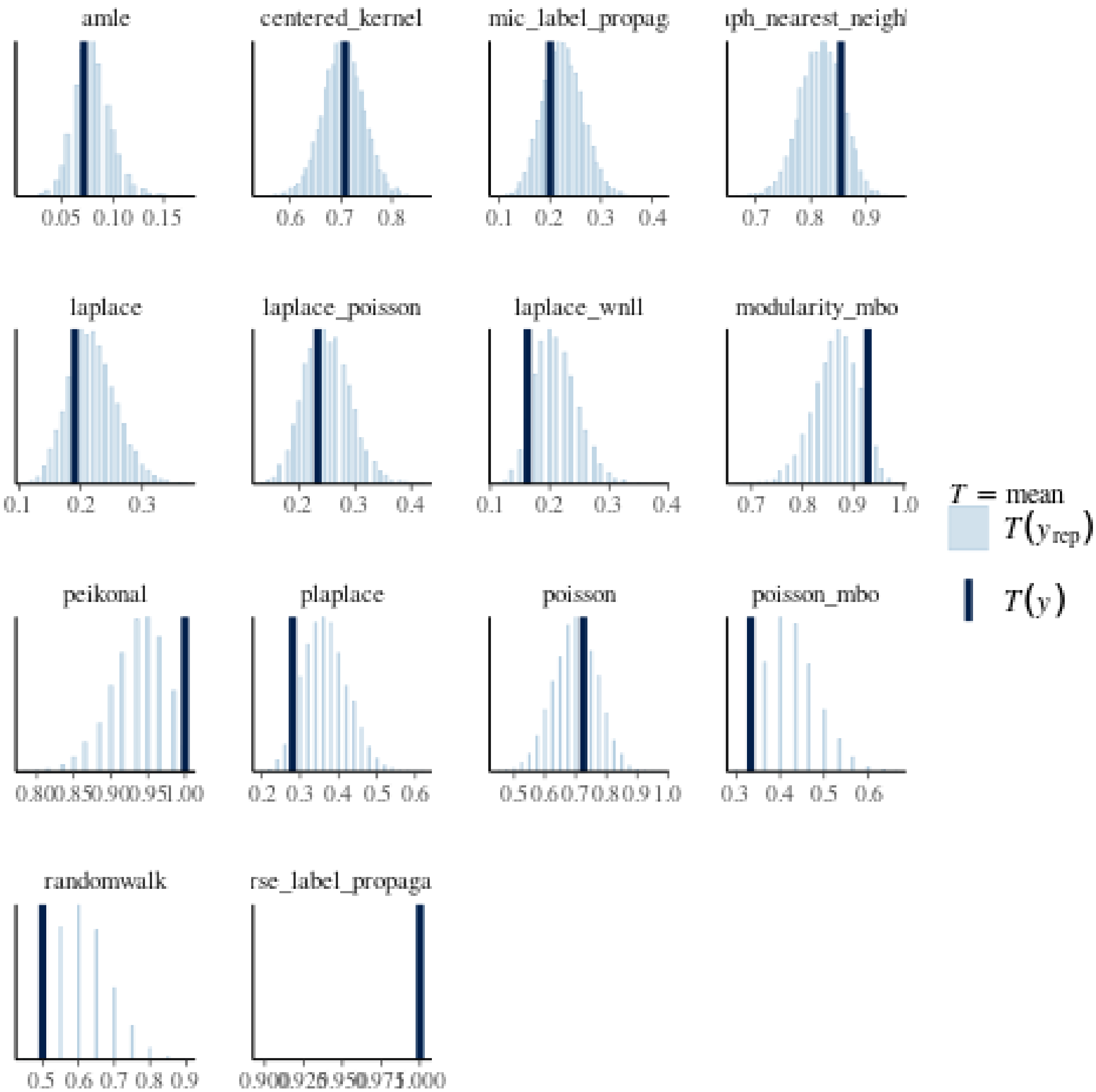

FIGURE 29. Checking posterior predictive test (75%) statistics within the region, in this case, the mean. The vertical lines are observed means.

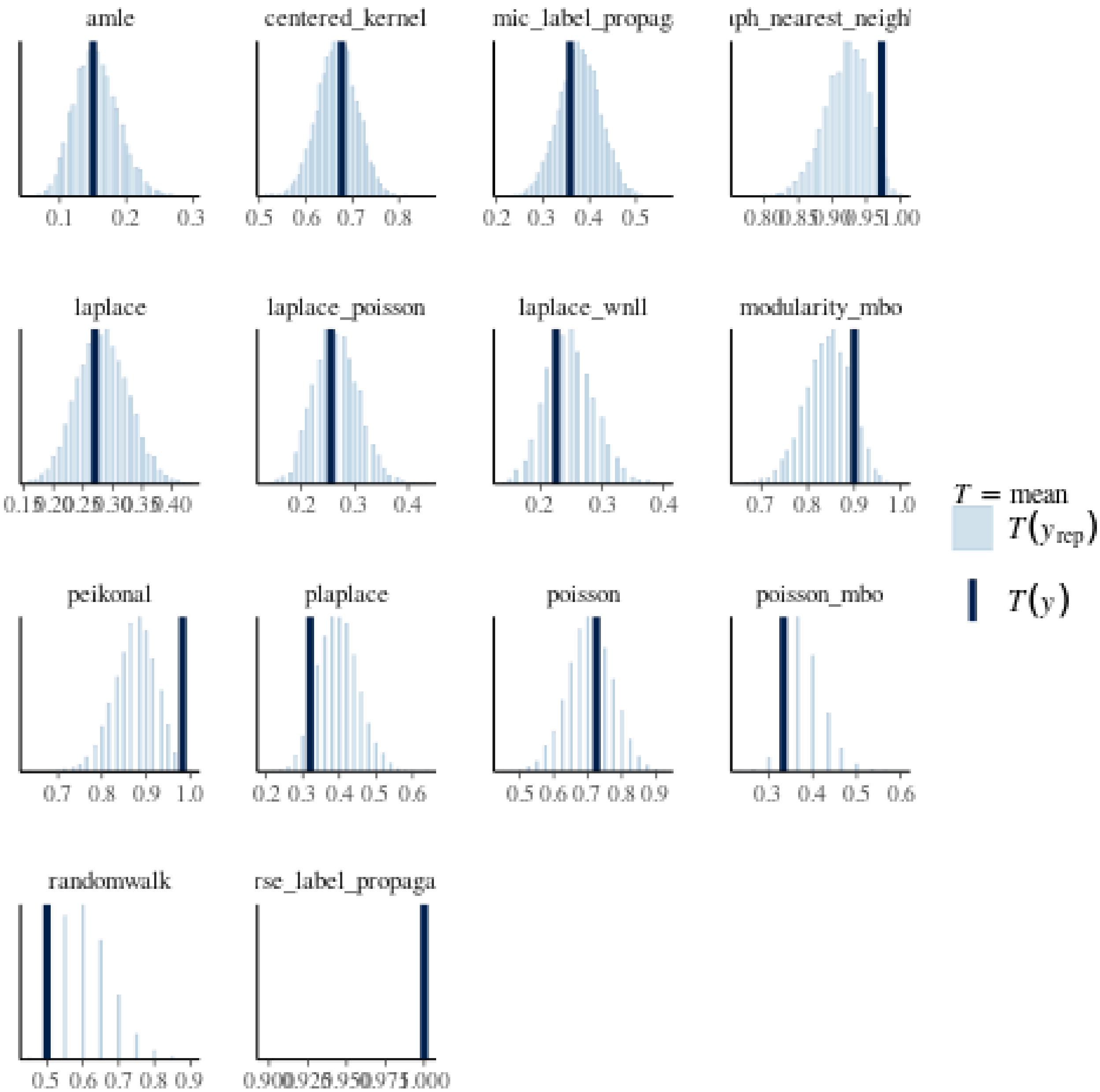

FIGURE 30. Checking posterior predictive test (90%)statistics within the region, in this case, the mean. The vertical lines are observed means.

References

[1] Jonah Gabry, Daniel Simpson, Aki Vehtari, Michael Betancourt, and Andrew Gelman. Visualization in bayesian workflow. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 182(2):389–402, 2019.

[2] David Issa Mattos, Jan Bosch, and Helena Holmström Olsson. Statistical models for the analysis of optimization algorithms with benchmark functions. *IEEE Transactions on Evolutionary Computation*, 25(6):1163–1177, 2021.

[3] Richard McElreath. *Statistical rethinking: A Bayesian course with examples in R and Stan.* CRC press, 2020.