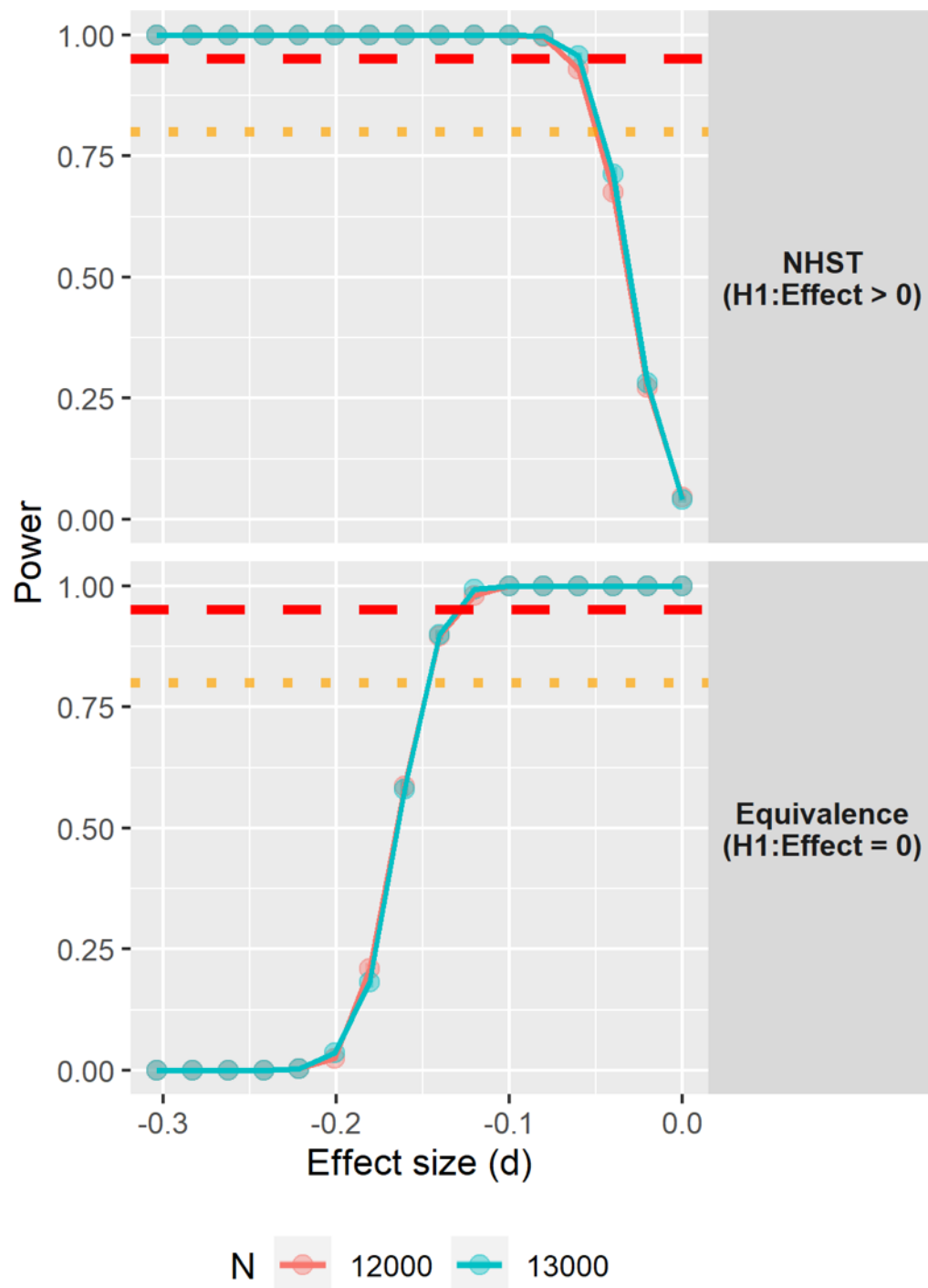
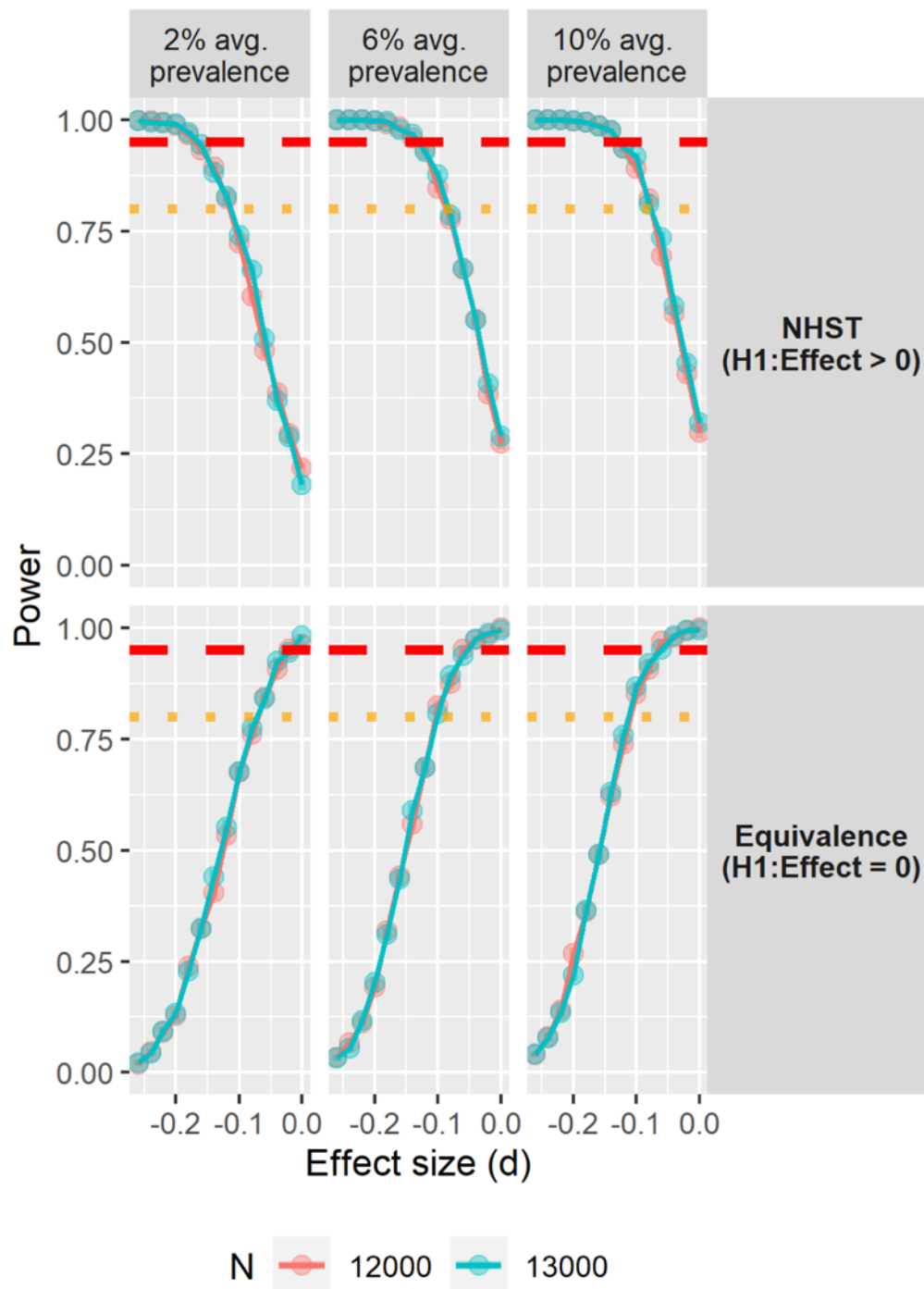


Additional file 8: Fig. S1. Power for NHST and equivalence test for H1a.



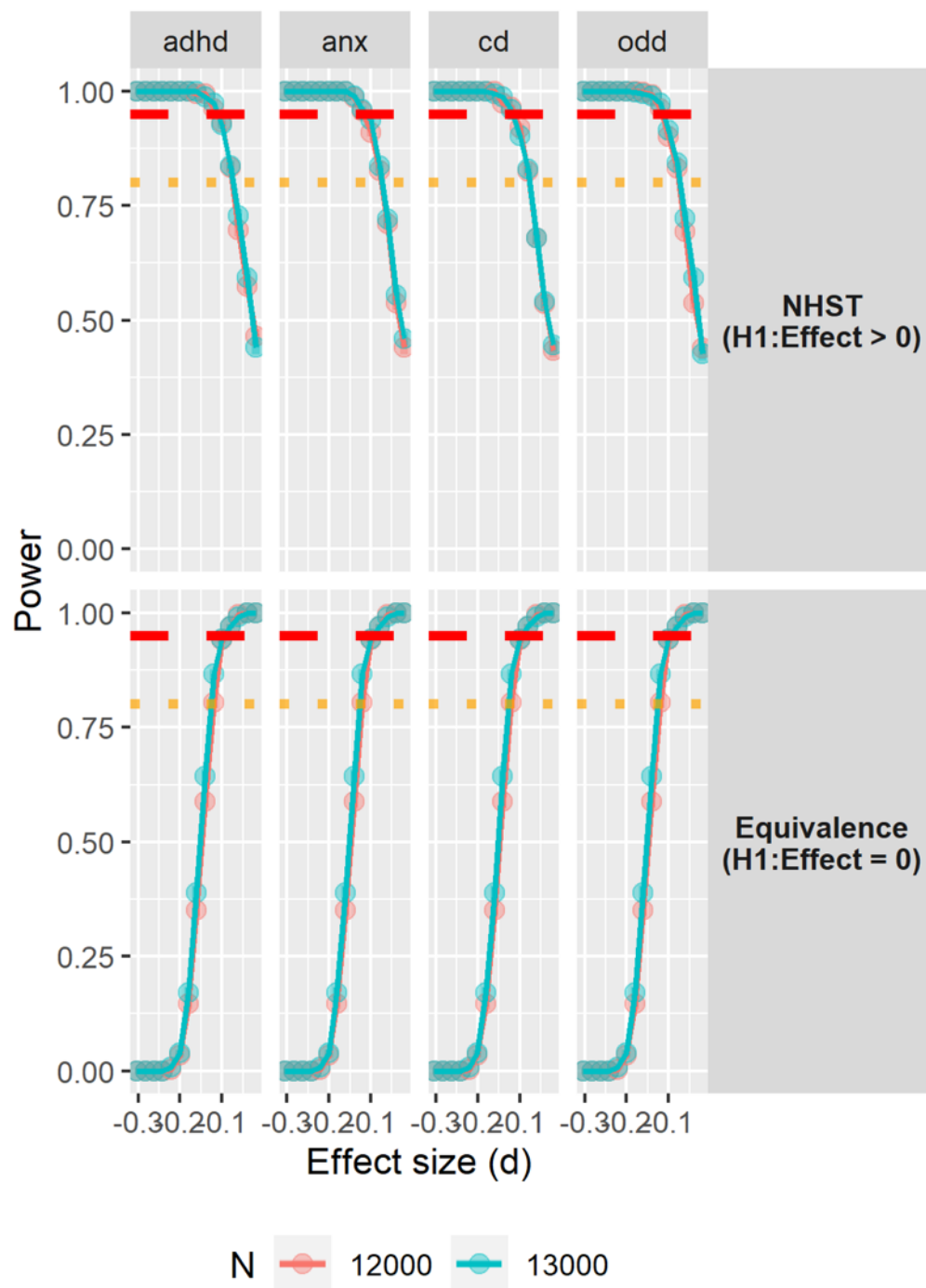
Power for testing hypothesis 1a at a range of potential effect sizes, with N=12,000 (light red) or N=13,000 (light green); dashed red line is the 95% power threshold, dotted yellow line is the 80% power threshold; NHST, null hypothesis significance testing.

Additional file 8: Fig. S2. Power for NHST and equivalence test for H1b.



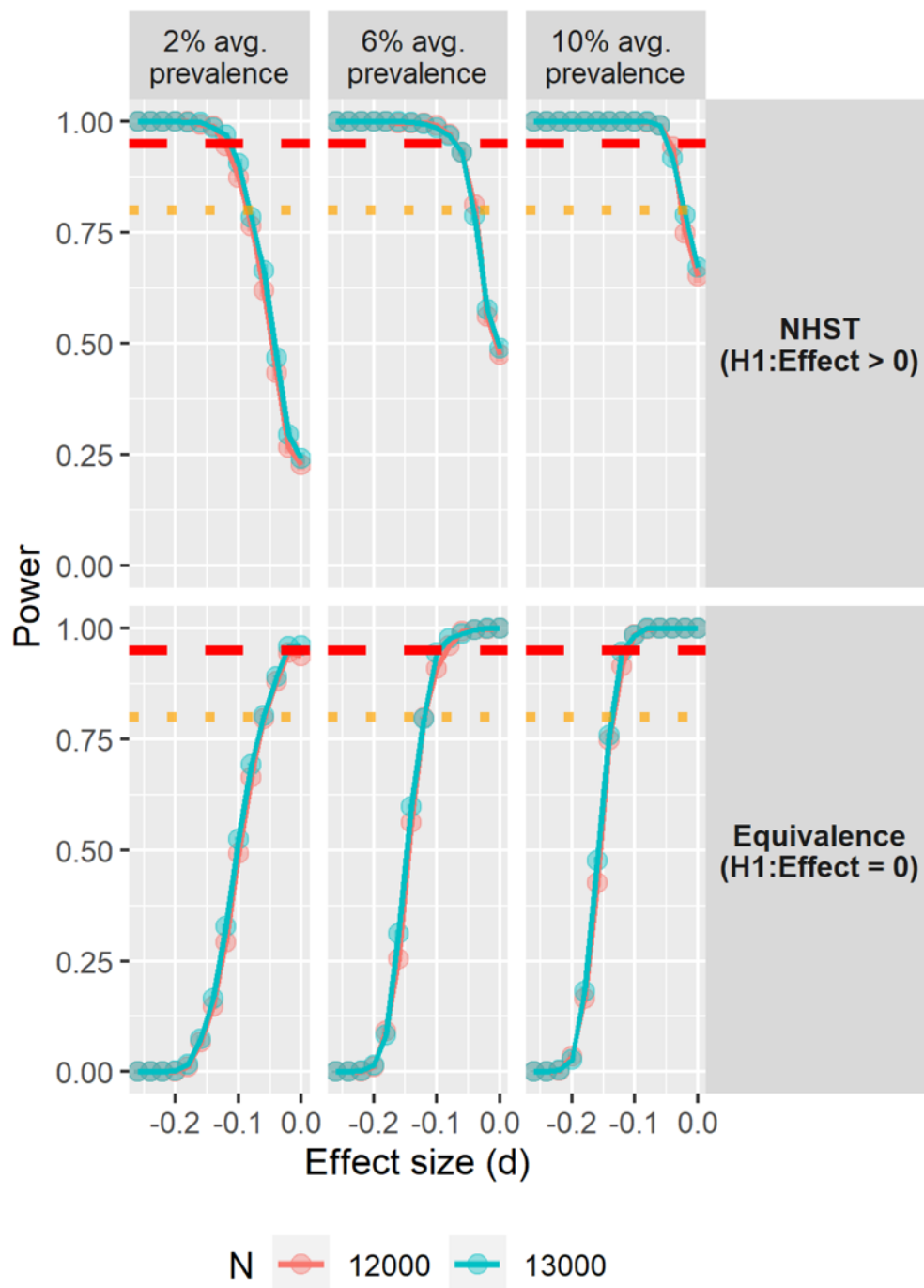
Power for testing hypothesis 1b at a range of potential effect sizes, with N=12,000 (light red) or N=13,000 (light green); dashed red line is the 95% power threshold, dotted yellow line is the 80% power threshold; NHST, null hypothesis significance testing.

Additional file 8: Fig. S3. Power for NHST and equivalence test for H2a.



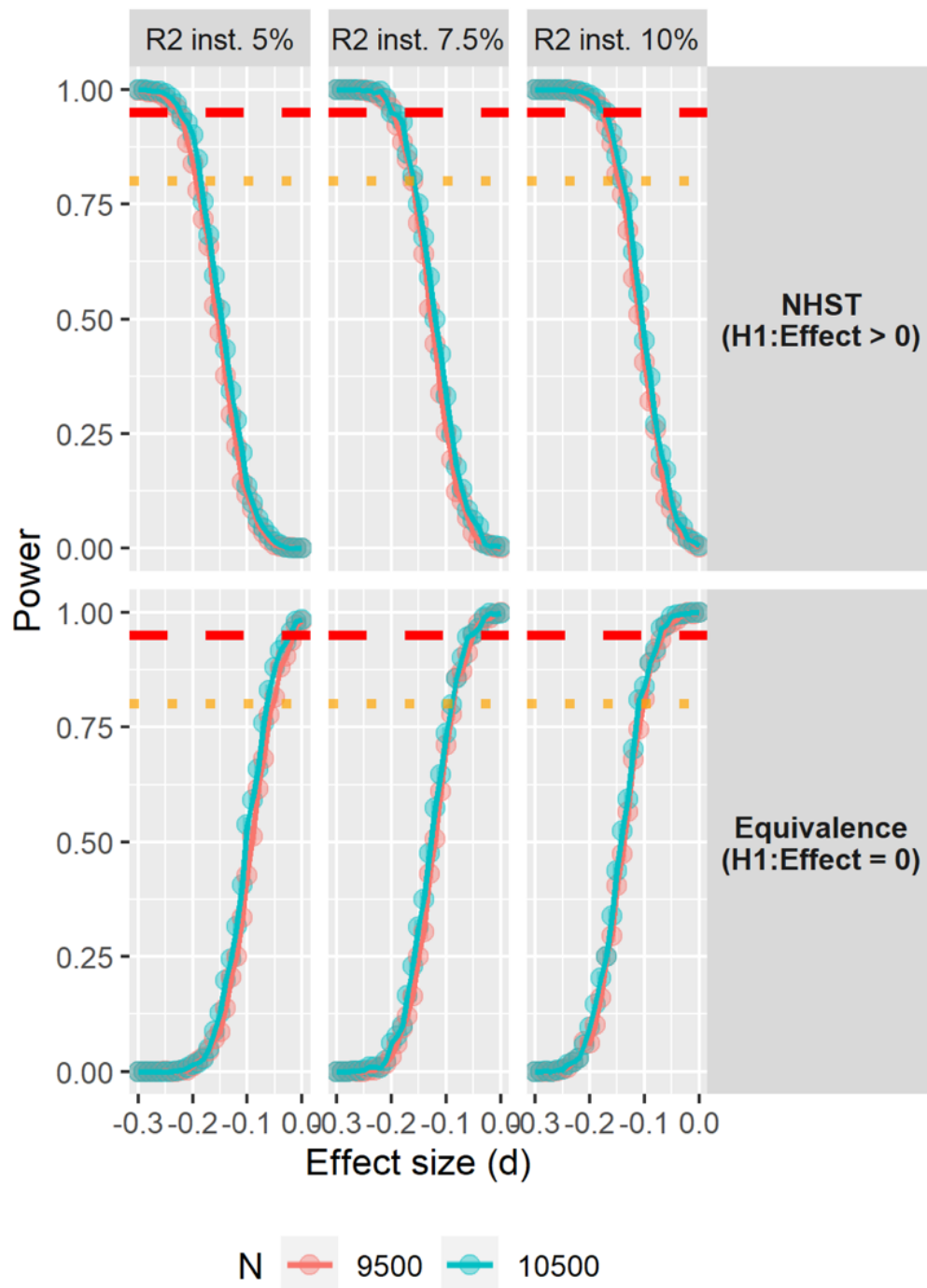
Power for testing hypothesis 2a at a range of effect sizes for each symptom domain, with N=12,000 (light red) or N=13,000 (light green); dashed red line is the 95% power threshold, dotted yellow line is the 80% power threshold; adhd, attention-deficit hyperactivity disorder; anx, anxiety; cd, conduct disorder; odd, oppositional defiant disorder; NHST, null hypothesis significance testing.

Additional file 8: Fig. S4. Power for NHST and equivalence test for H2b.



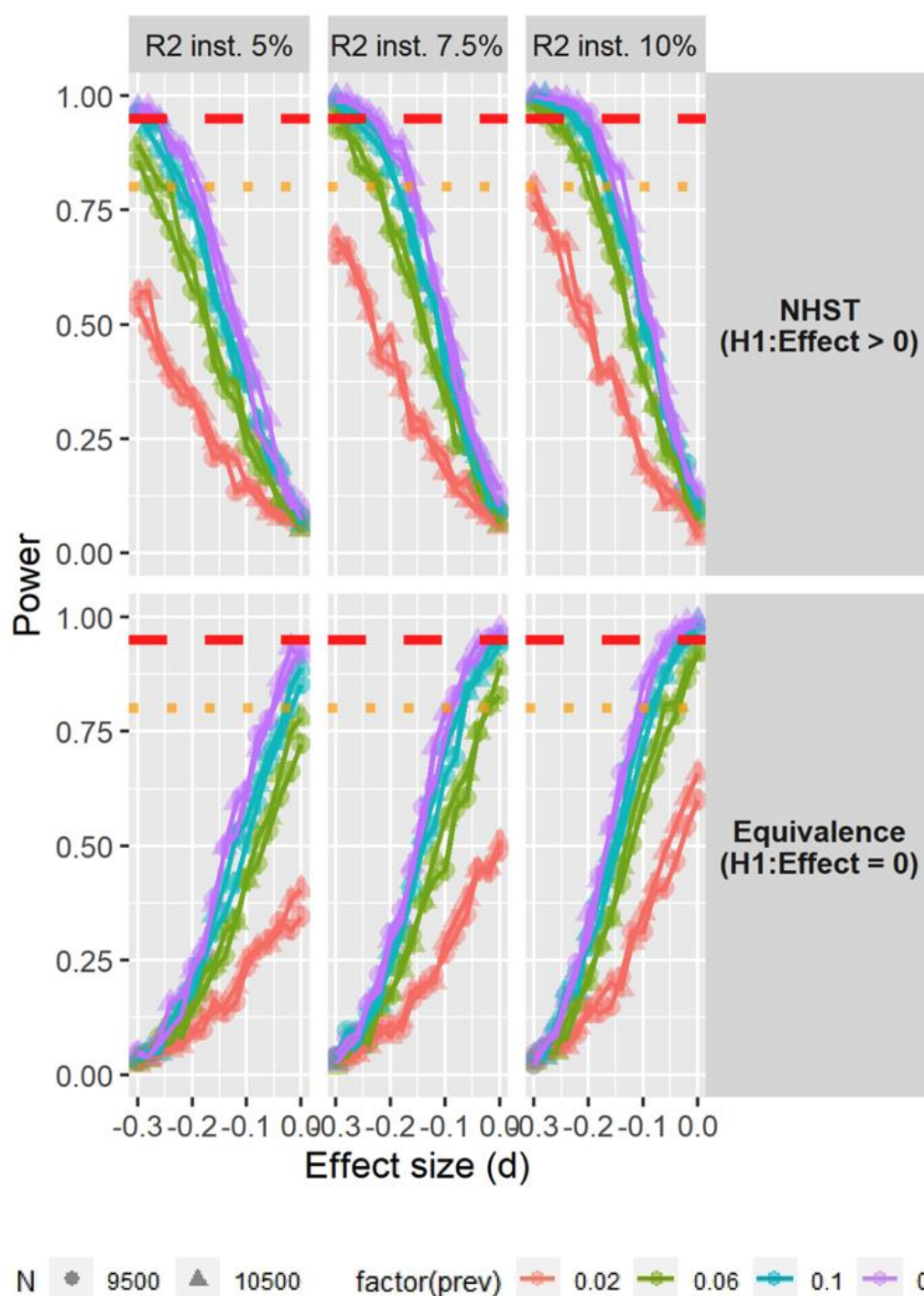
Power for testing hypothesis 2b at a range of effect sizes for 2%, 6%, and 10% avg. prevalence of a disorder, with $N=12,000$ (light red) or $N=13,000$ (light green); dashed red line is the 95% power threshold, dotted yellow line is the 80% power threshold; NHST, null hypothesis significance testing.

Additional file 8: Fig. S5. Power for NHST and equivalence test for H3a.



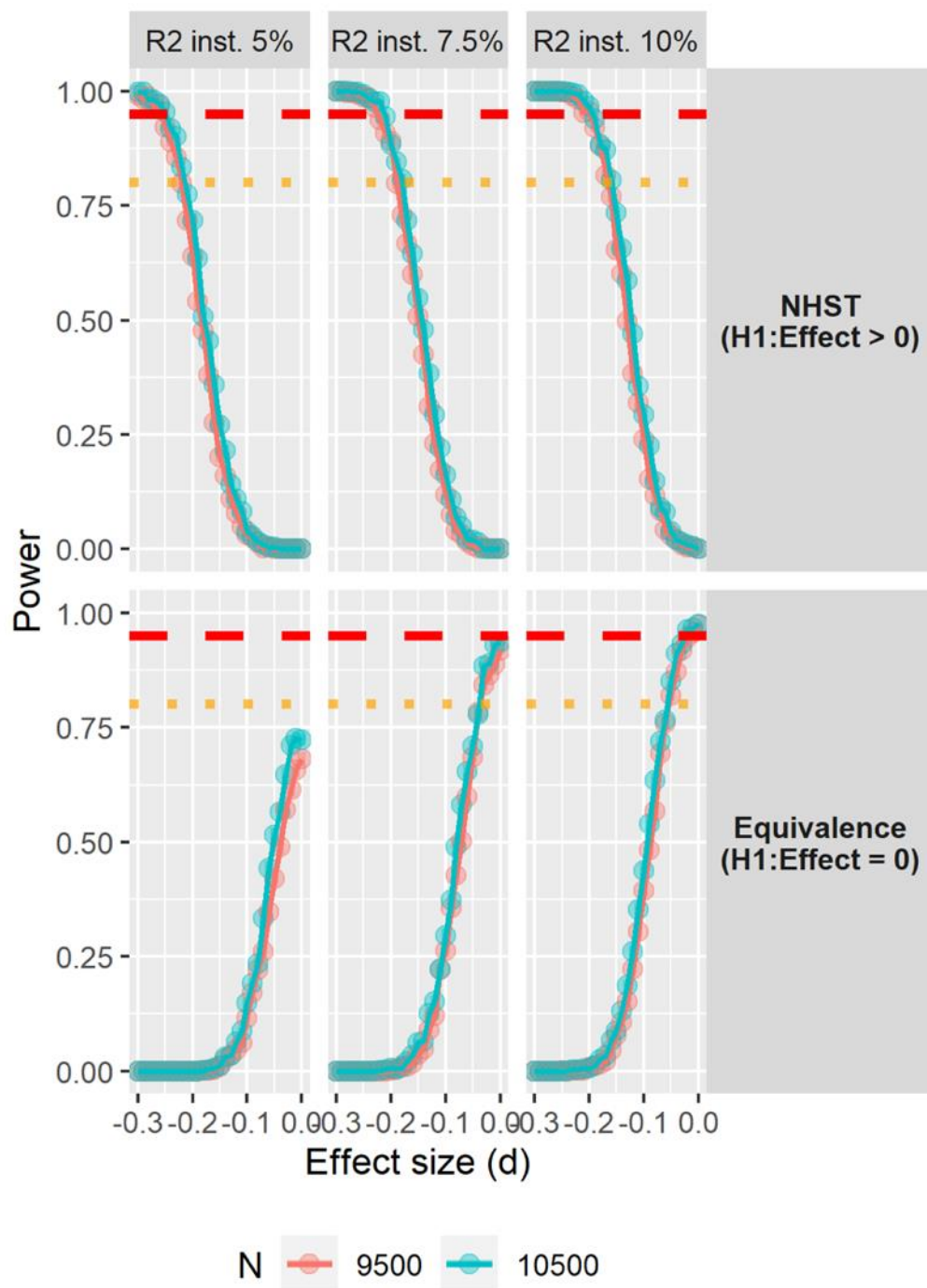
Power for testing hypothesis 3a at a range of effect sizes with 5%, 7.5%, and 10% instrument strength, at $N=9,500$ (light red) or $N=10,500$ (light green); dashed red line is the 95% power threshold, dotted yellow line is the 80% power threshold; R^2 inst, R^2 instrument-exposure association; NHST, null hypothesis significance testing.

Additional file 8: Fig. S6. Power for NHST and equivalence test for H3b.



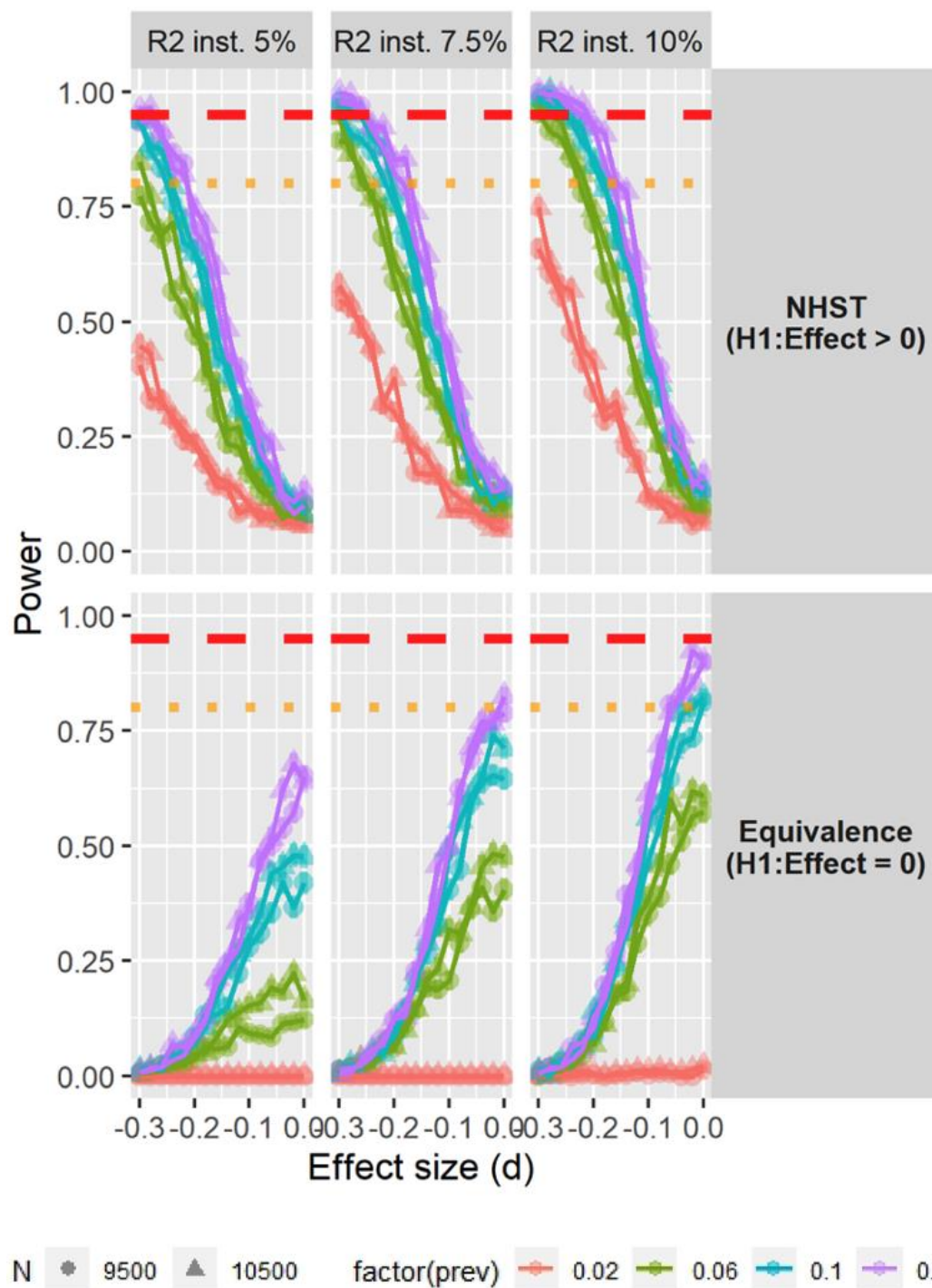
Power for testing hypothesis 3b at a range of effect sizes with 5%, 7.5%, and 10% instrument strength, across four prevalence rates (2% in light red, 6% in green, 10% in light blue, and 14% in purple), at N=9,500 (points) or N=10,500 (triangles); dashed red line is the 95% power threshold, dotted yellow line is the 80% power threshold; R2 inst, R² instrument-exposure association; NHST, null hypothesis significance testing.

Additional file 8: Fig. S7. Power for NHST and equivalence test for H4a.



Power for testing hypothesis 4a at a range of effect sizes with 5%, 7.5%, and 10% instrument strength, with N=9,500 (light red) or N=10,500 (light green); dashed red line is the 95% power threshold, dotted yellow line is the 80% power threshold; R2 inst, R² instrument-exposure association; NHST, null hypothesis significance testing.

Additional file 8: Fig. S8. Power for NHST and equivalence test for H4b.



Power for testing hypothesis 4b at a range of effect sizes with 5%, 7.5%, and 10% instrument strength, across four prevalence rates (2% in light red, 6% in green, 10% in light blue, and 14% in purple), at $N=9,500$ (points) or $N=10,500$ (triangles); dashed red line is the 95% power threshold, dotted yellow line is the 80% power threshold; R2 inst, R^2 instrument-exposure association; NHST, null hypothesis significance testing.