

Always Use the Separate Variances t Test for Two Independent Groups

Joshua D. Wondra and Richard Gonzalez

University of Michigan

Abstract

This is an abstract

Keywords: t test, new statistics

Always Use the Separate Variances t Test for Two Independent Groups

This is an introduction

Method

Results

We report the observed Type I error rates, observed power, and coverage probability for the classic and separate variance t tests.

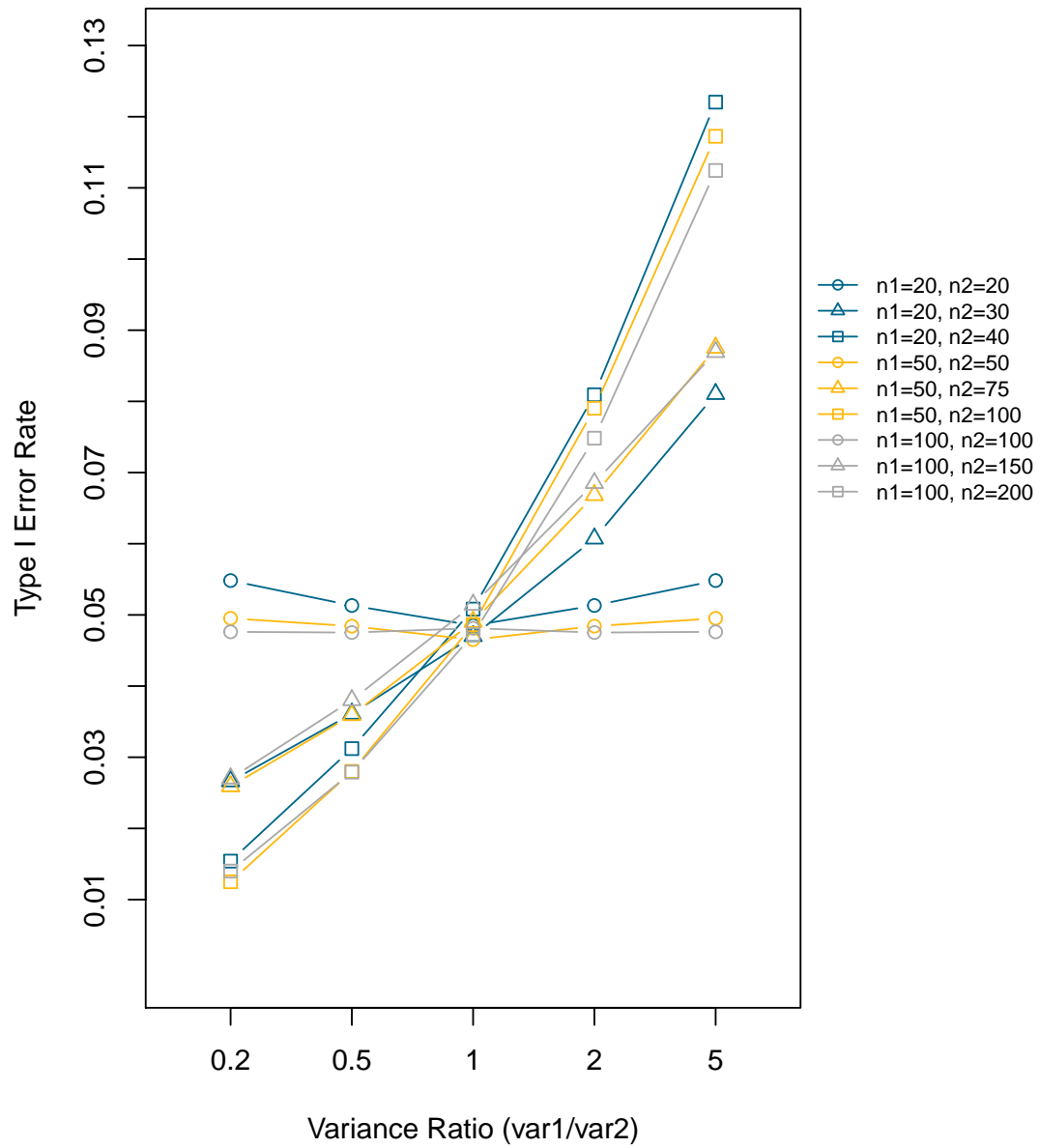
Type I Error Rates

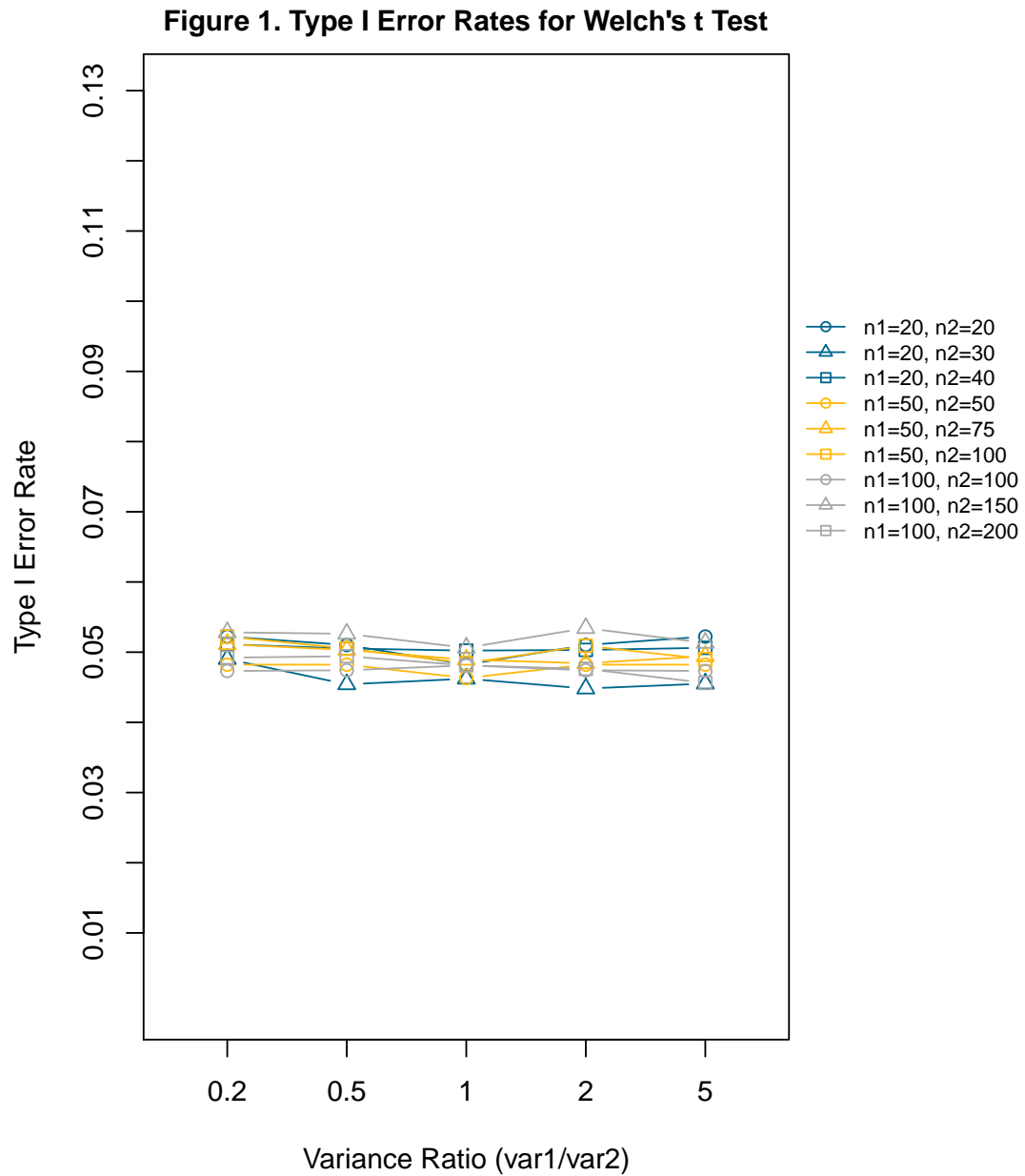
In this section we report type I error rates for the classic and separate variance t tests when the null hypothesis is true. Prior research has demonstrated that when either sample sizes or variances are equal, the type I error rates are preserved at .05 for both tests (CITATIONS).

Equal Variances, Different Sample Sizes. Consistent with prior research, the type I error rate was .05 for both the classic and separate variance t tests across all sample sizes and sample size ratios.

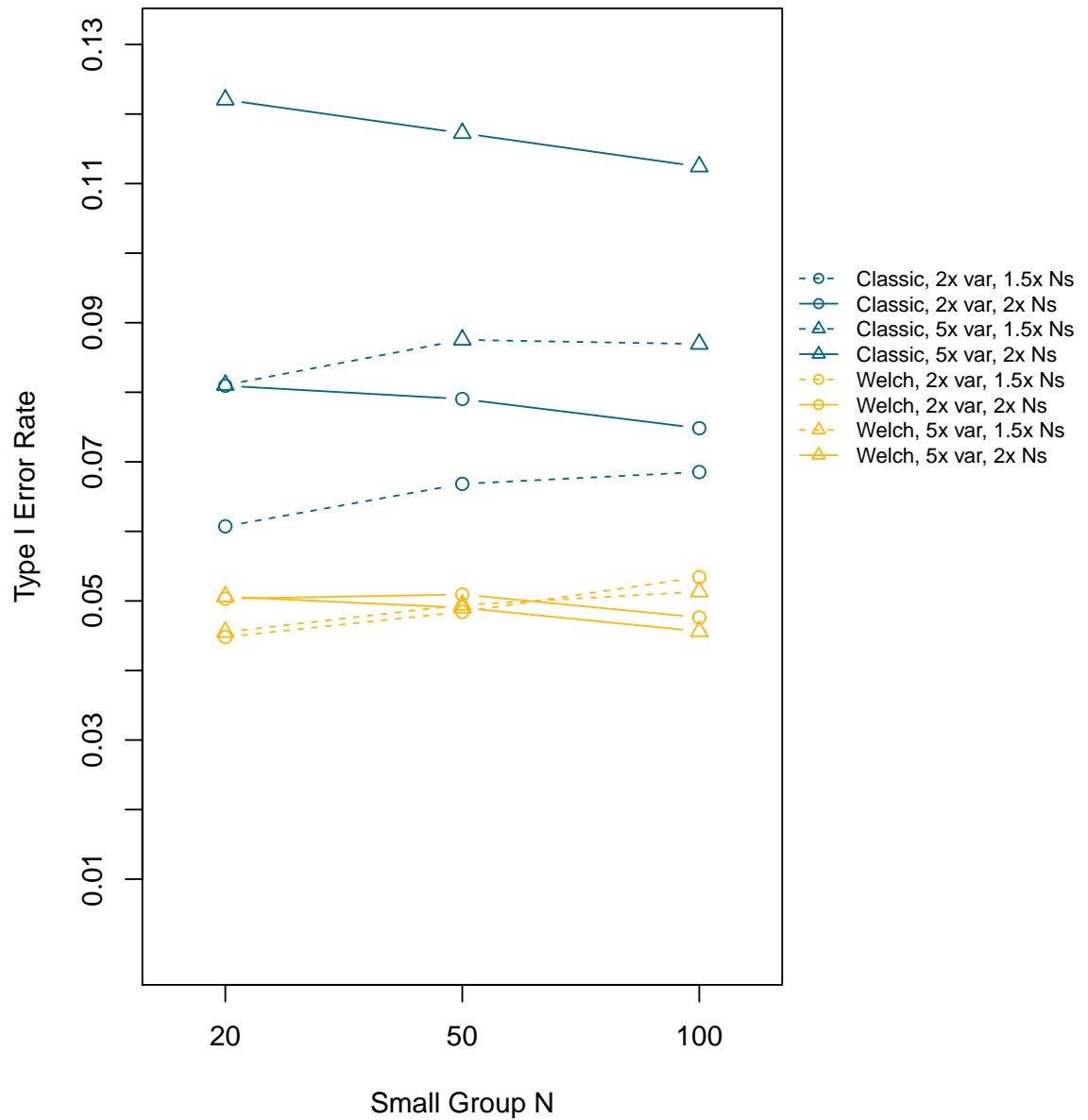
Equal Sample Sizes, Different Variances. Consistent with prior research, the type I error rate was .05 for both the classic and separate variance t tests across all sample sizes and variance ratios.

Different Sample Sizes, Different Variances. Consistent with prior research, the classic t test was not robust when both variances and sample sizes were equal. When the group with the bigger sample also had the bigger variance, the classic t test was more conservative; the type I error rate dropped as low as .01 when the large group had twice the sample size and five times the variance of the small group (see Figure 1).

Figure 1. Type I Error Rates for Student's t Test



When the group with the bigger sample had the smaller variance, however, the classic t test was more liberal; the type I error rate inflated to just over .12 when the large group had twice the sample size and five times the variance of the small group (see Figure 2). In all cases, however, the separate variances t test approximately maintained the expected type I error rate of .05.

Figure 2. Type I Error Rates When the Bigger Sample Has Smaller Variance

In summary, the type I error rates for the classic t test were quite variable when both sample sizes and variances were unequal between the two groups. In contrast, the type I error rates for the separate variances t test were mostly consistent with the expected .05 rate.

Power

In this section we report observed power rates for the classic and separate variance t tests when the null hypothesis is true. Prior research has focused on Type I error rates, but it has not examined whether choosing the separate variance t test also results in a loss of power to detect true effects. We examined power differences in the two tests when the sample sizes and variances are unequal. We looked at the power of the tests to detect small, medium, and large effects based on Cohen's d and compared it to the expected power based on an assumption of equal variances. This means that for the baseline comparison when variances were unequal, we treated it as though the difference in variance is due to sampling error rather than a true population difference. As with Type I errors, there were no differences between the classic and separate variance tests when sample sizes and variances were equal.

Equal Variances, Different Sample Sizes. Figure 3 displays the observed power for the classic and separate variance tests when sample sizes and sample size ratios varied, but variances were equal. As with Type I errors, the differences in the proportion of rejected null hypotheses was negligible.

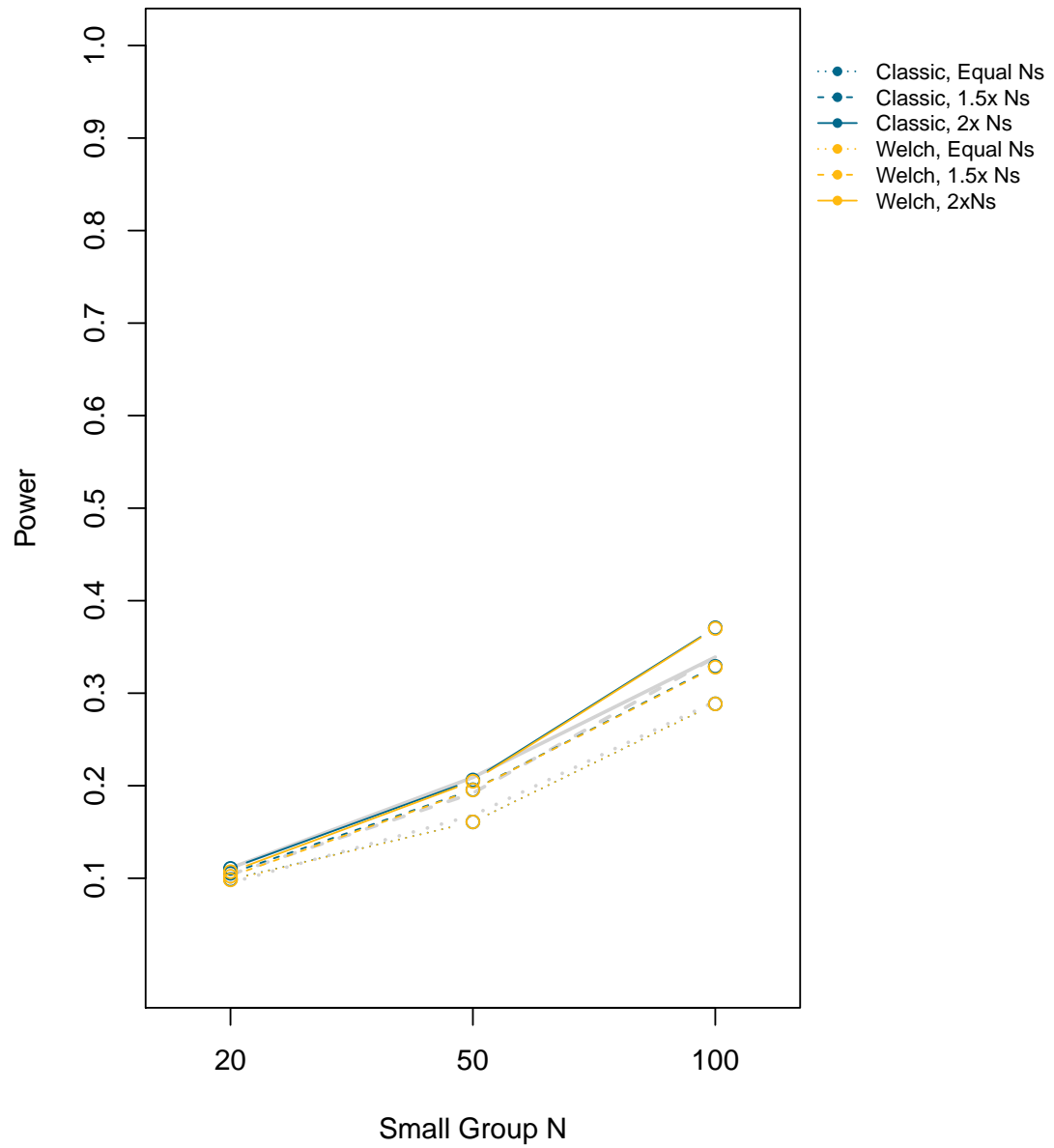
Figure 3. Power for Equal Variances and Different Sample Sizes with a Small Effect

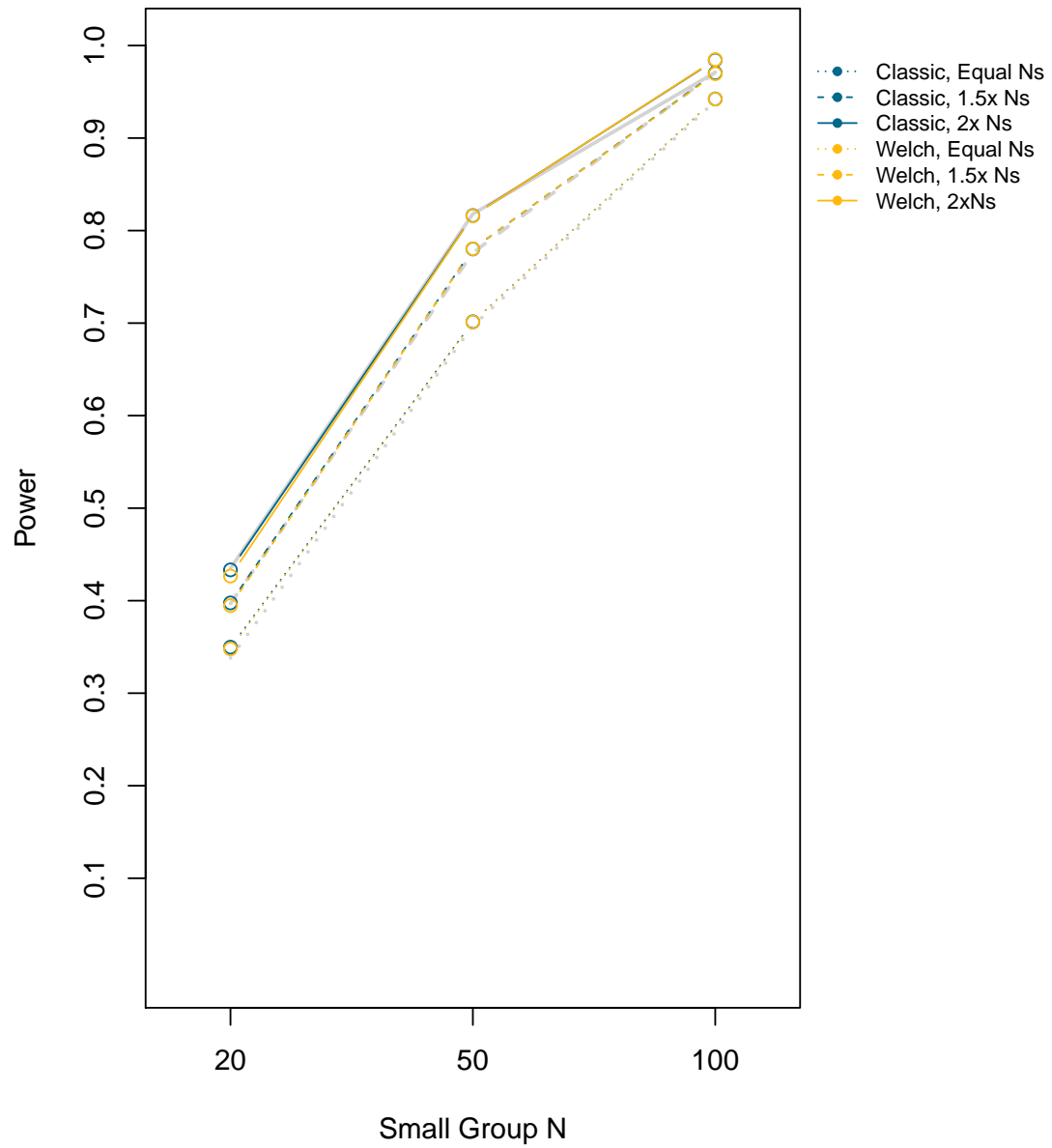
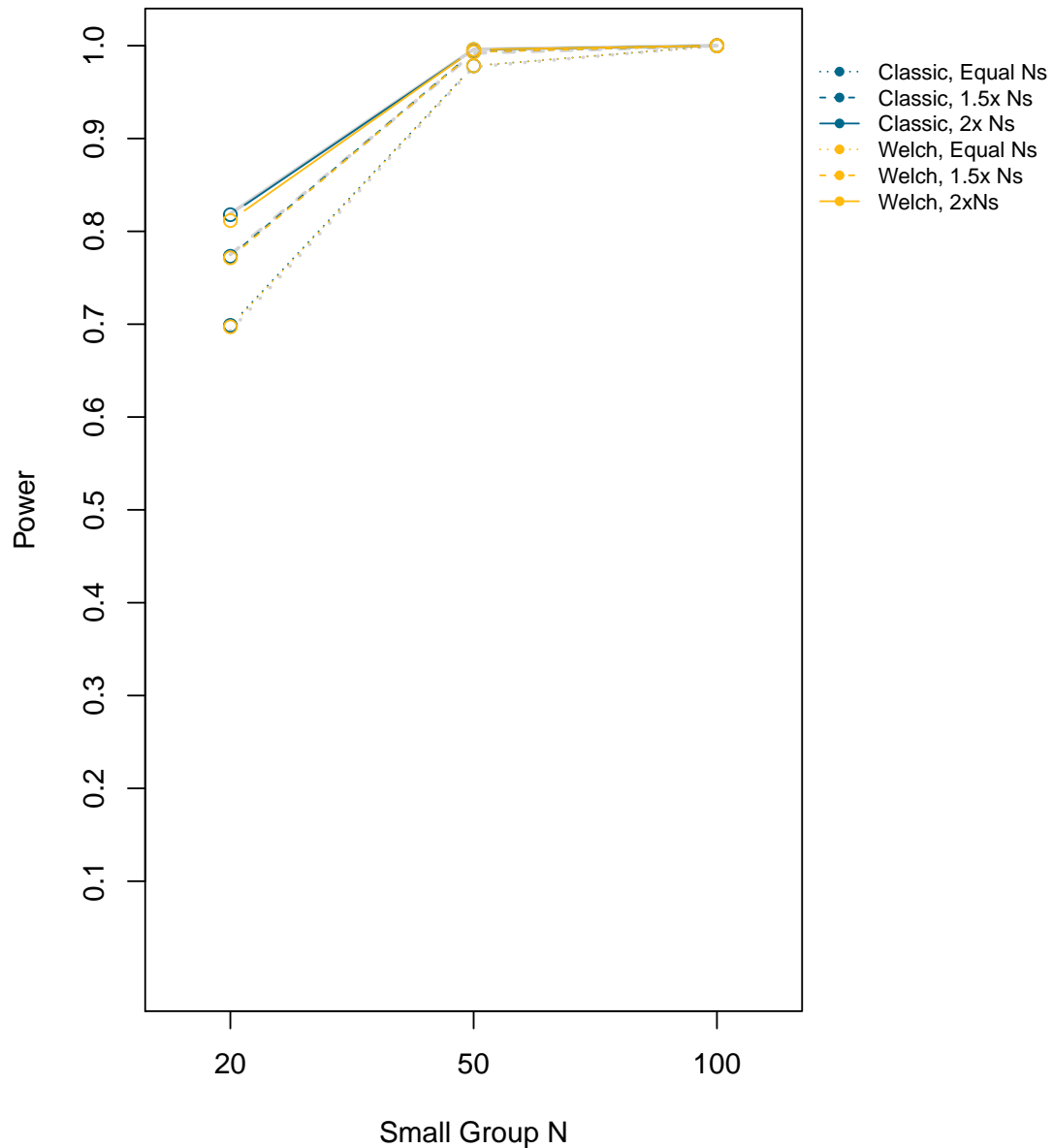
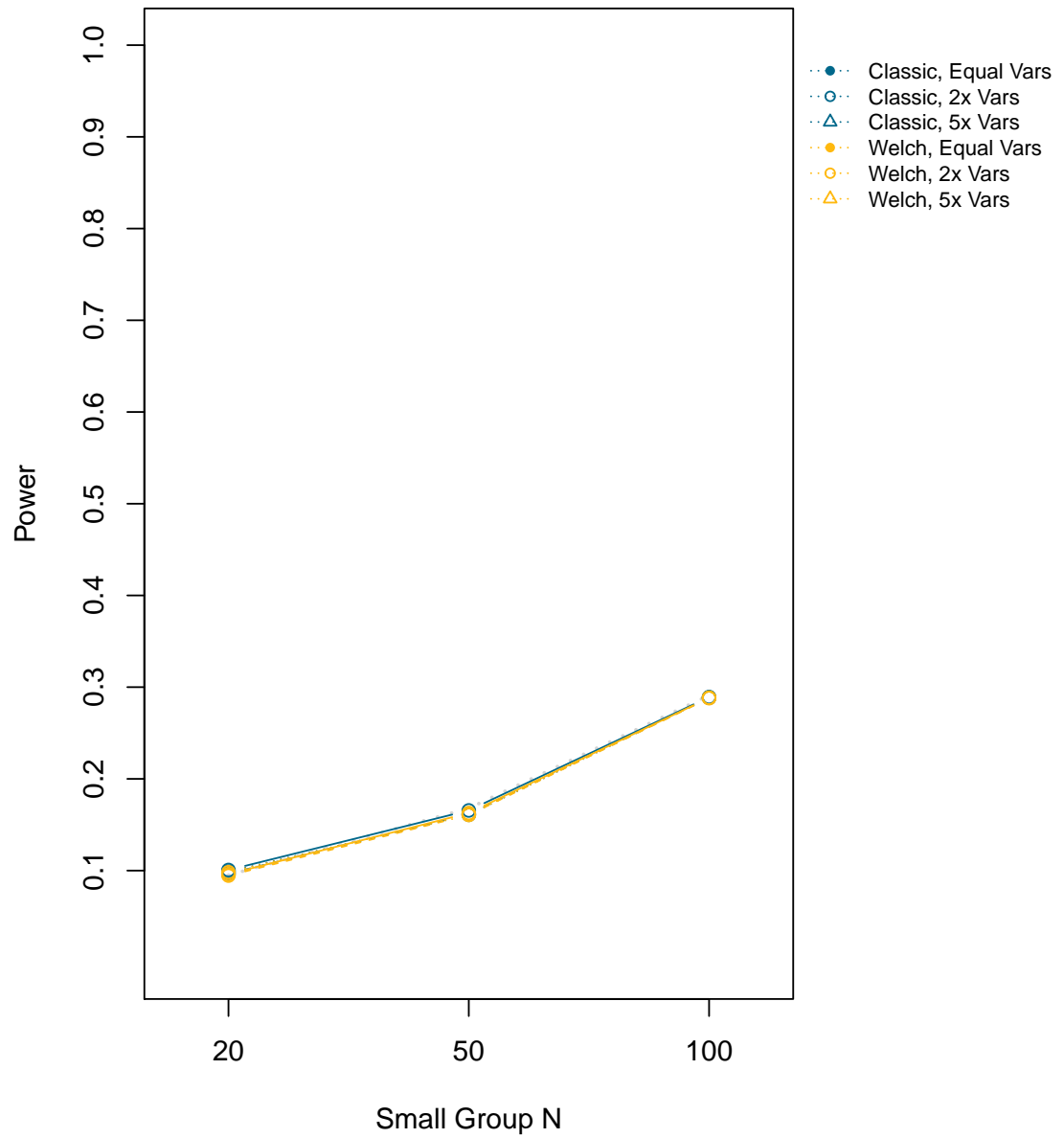
Figure 4. Power for Equal Variances and Different Sample Sizes with a Medium Effect

Figure 5. Power for Equal Variances and Different Sample Sizes with a Large Effect

Equal Sample Sizes, Different Variances. Figures 6-8 display the observed power for the classic and separate variance tests when sample sizes and sample size ratios are equal, but variance ratios vary. Here too there were essentially no differences in the proportion of rejected null hypotheses (NOTE: The mean differences between groups differ by the variance ratio and power is based on a standardized effect size, not an absolute mean difference).

Figure 6. Power for Unequal Variances and Different Sample Sizes with a Small Effect

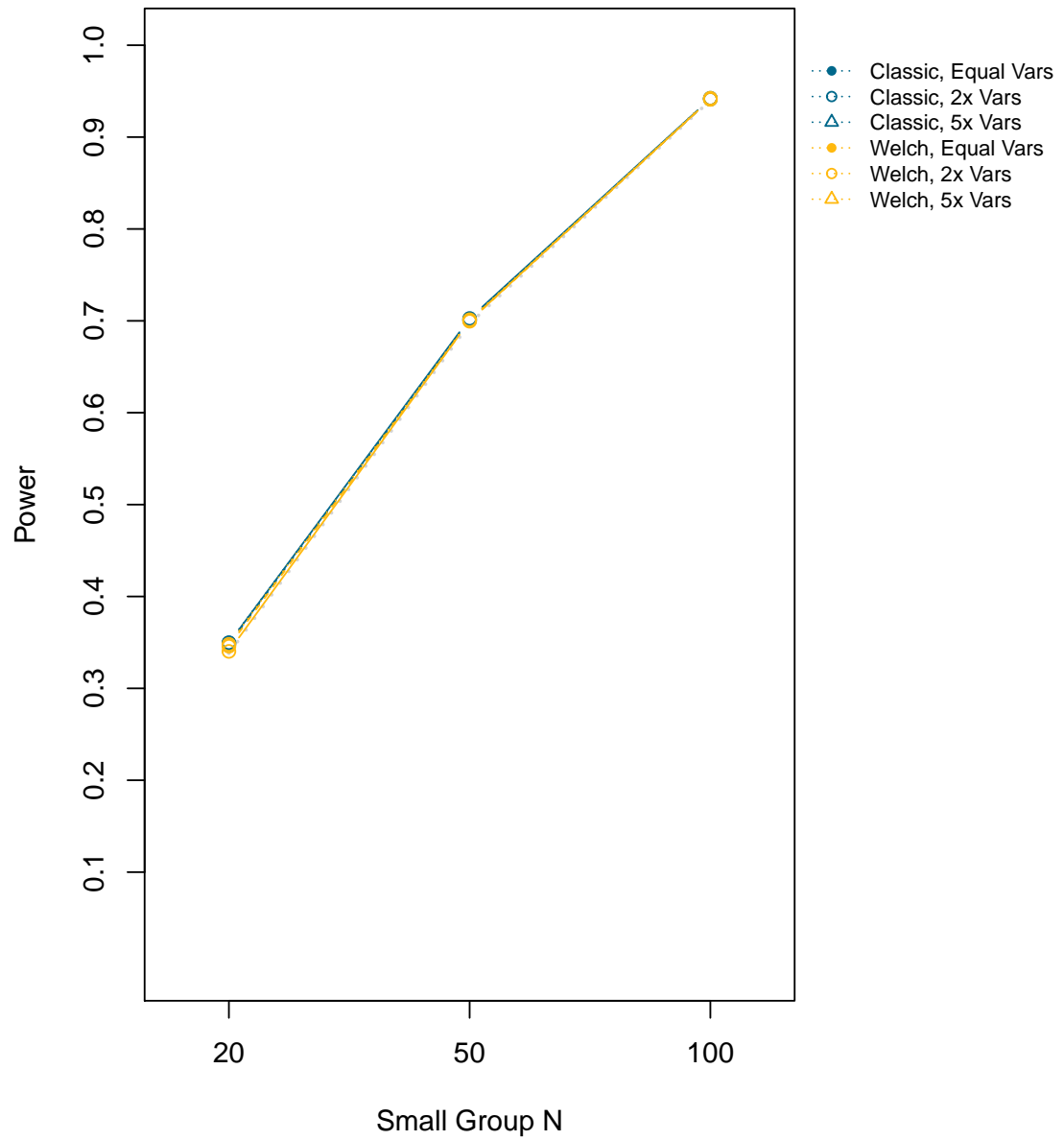
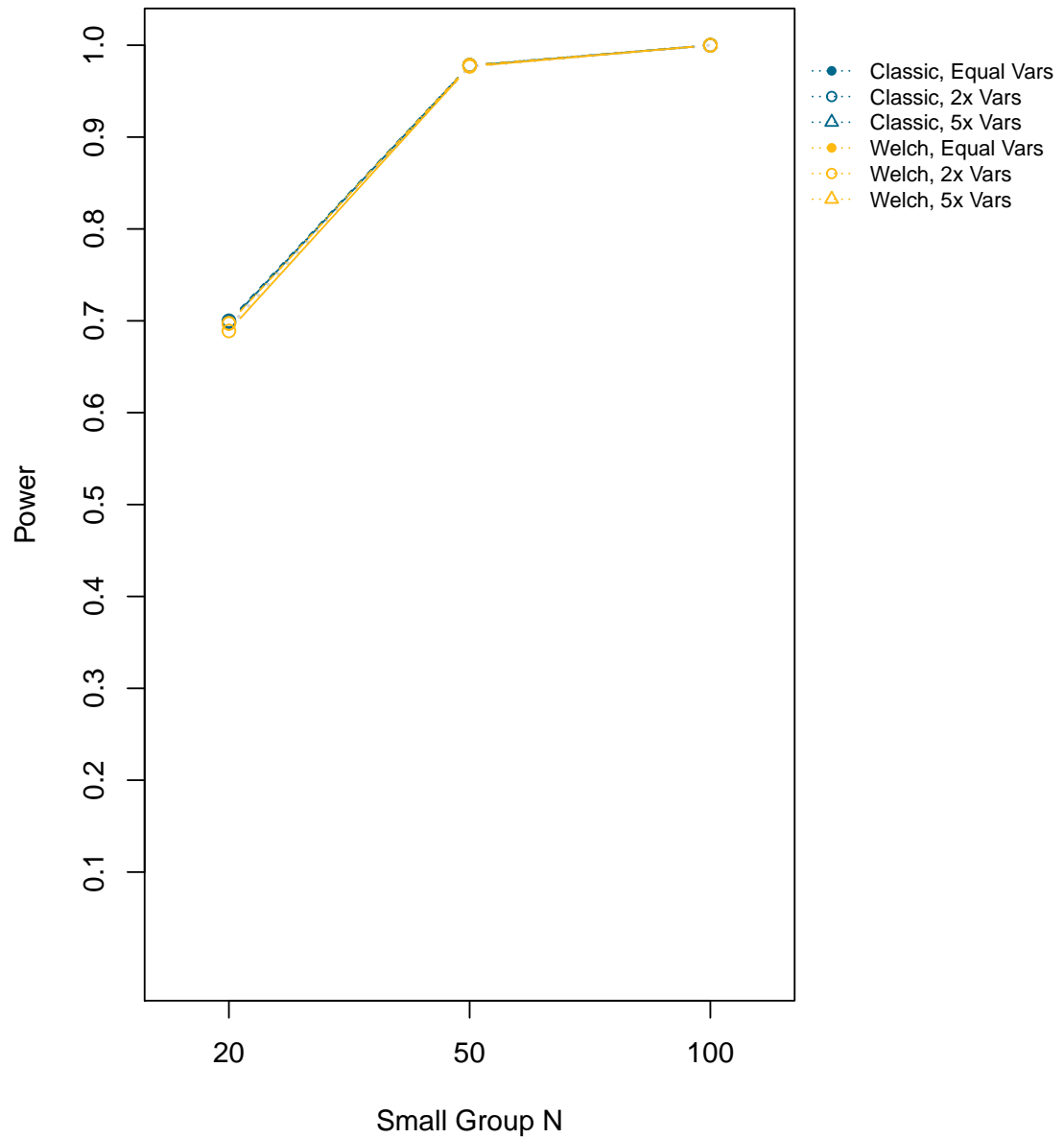
re 7. Power for Unequal Variances and Different Sample Sizes with a Medium Effect

Figure 8. Power for Unequal Variances and Different Sample Sizes with a Large Effect

Different Variances, Different Ns. The following tables display power when both variances and Ns differ.

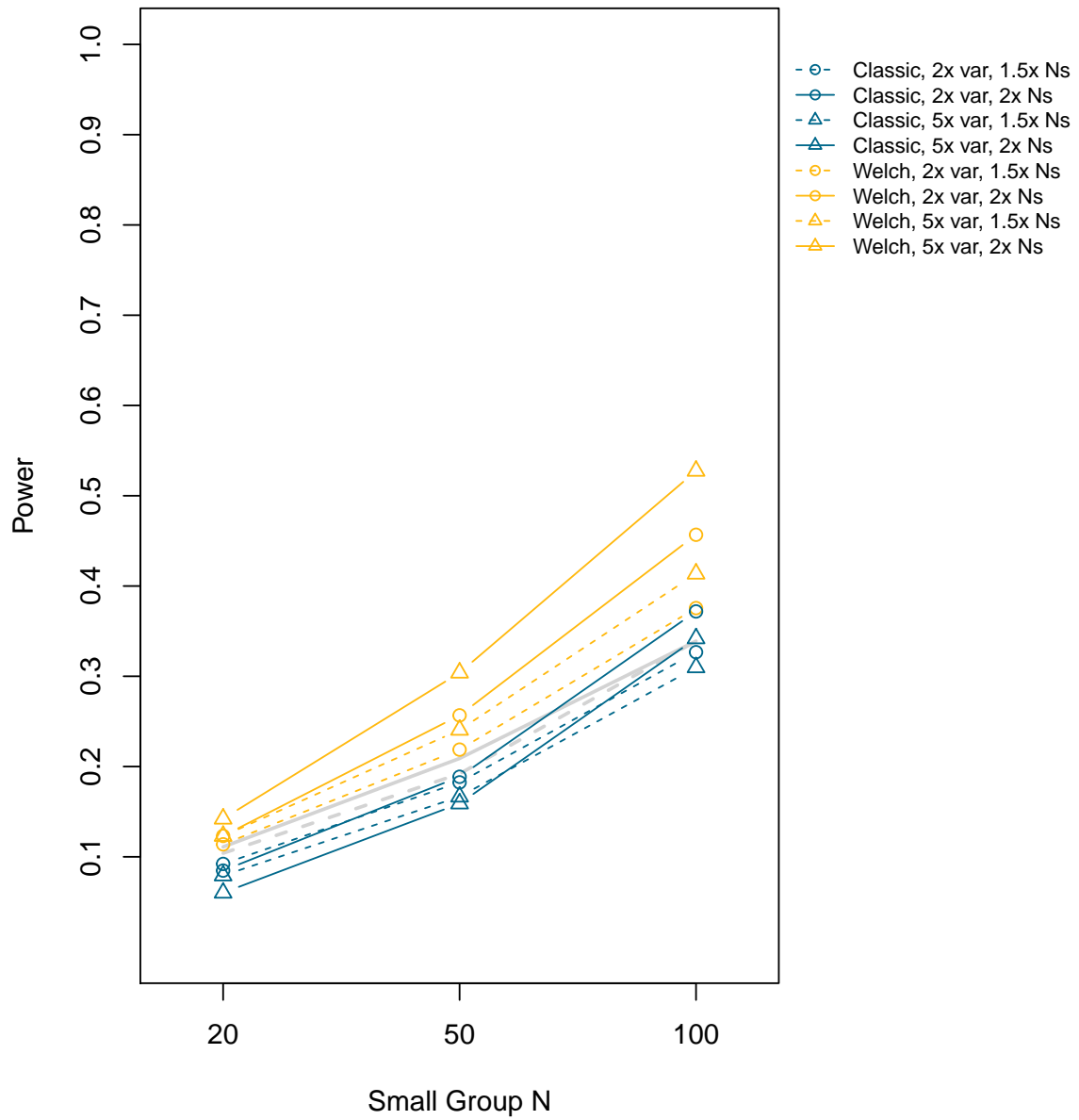
Figure 9. Power When the Bigger Sample Has Bigger Variance, Small Effect

Figure 9. Power When the Bigger Sample Has Bigger Variance, Medium Effect

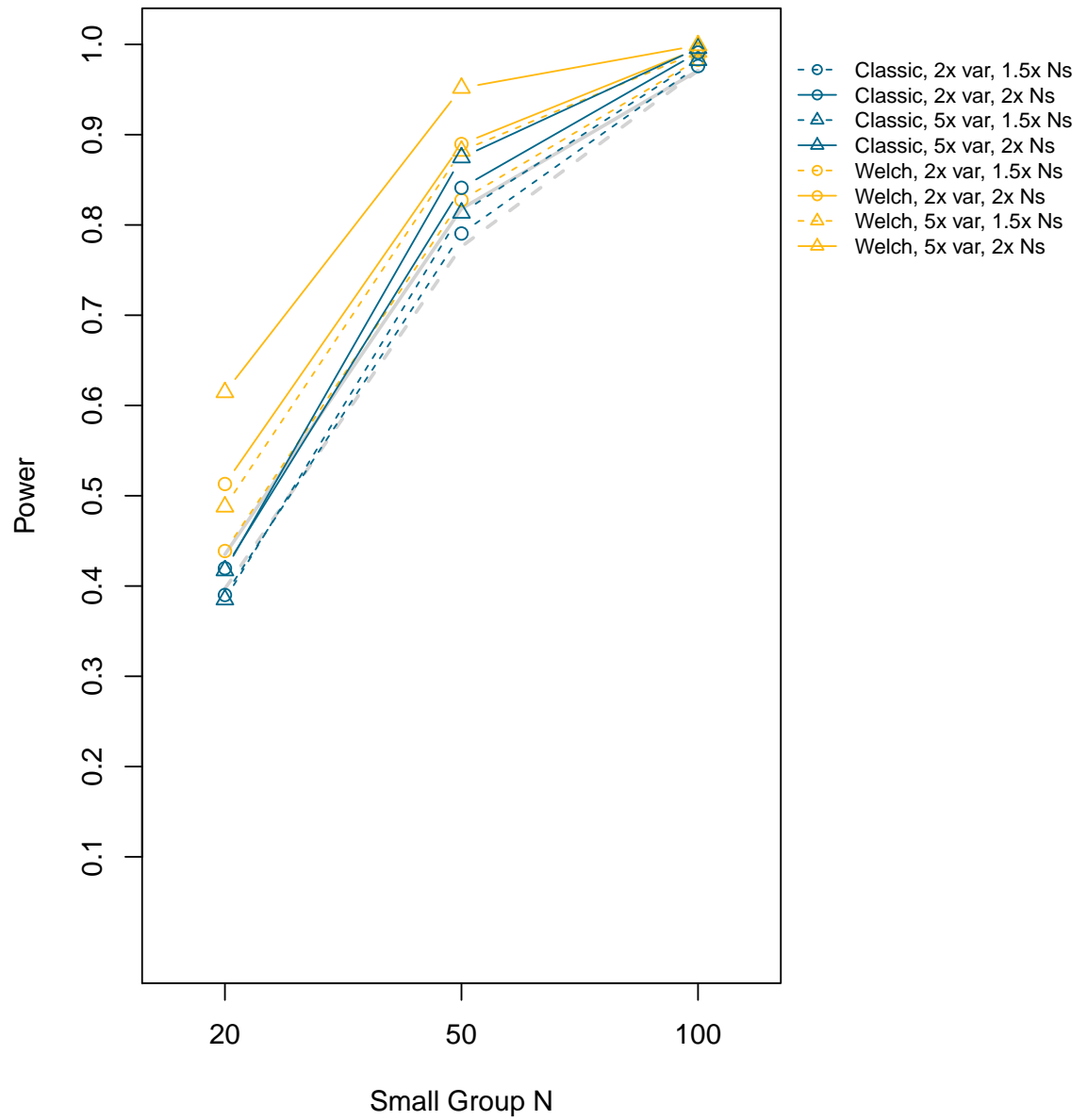
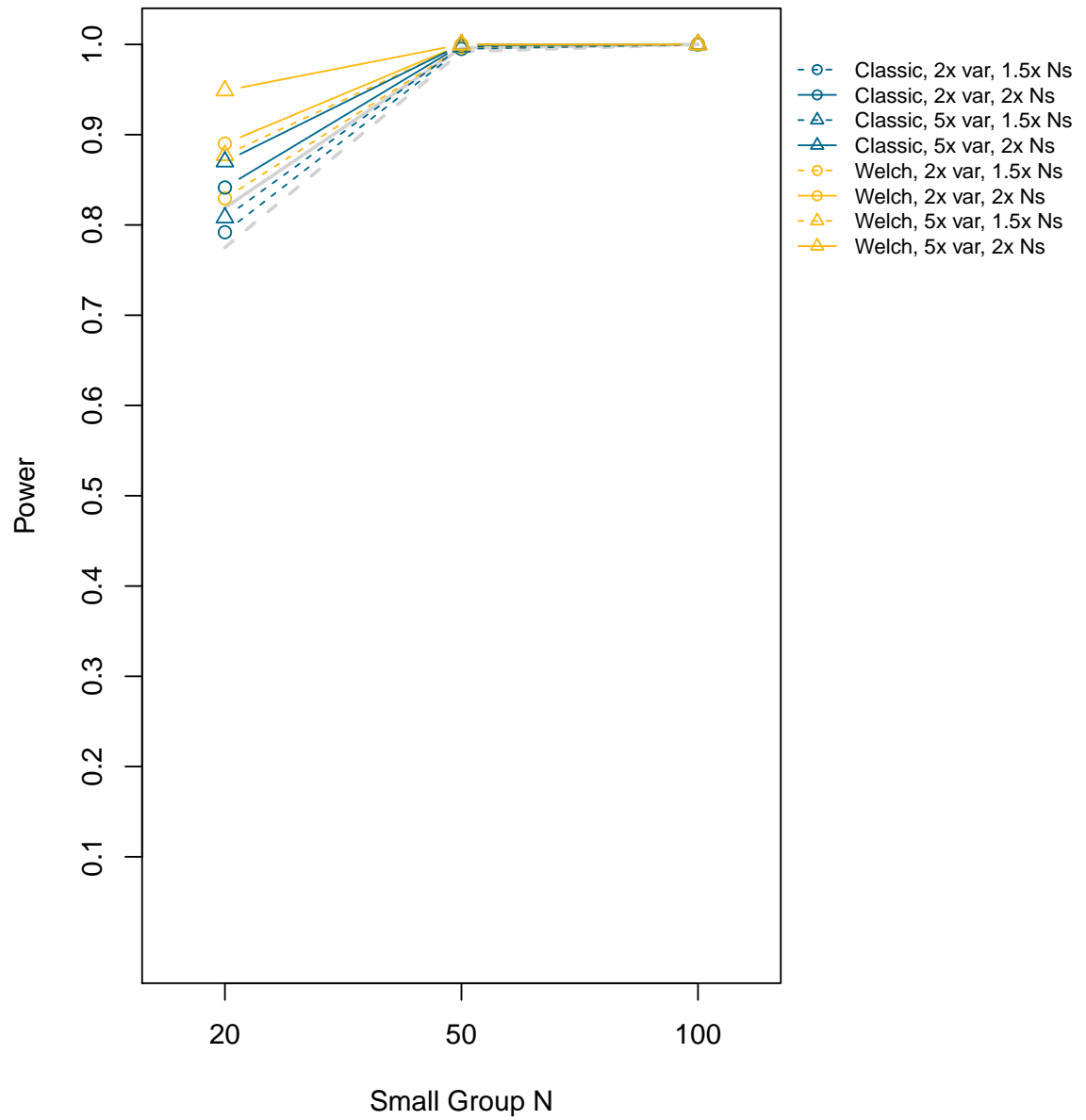


Figure 9. Power When the Bigger Sample Has Bigger Variance, Large Effect

Coverage Probability