and Welch's *t*-test. Even with extremely large SDR (respectively, 0.01, 0.1, 10, and 100) and small sample sizes (10 subjects per group), the biggest increase in power of Student's *t*-test compared to Welch's-*t* test is approximately 5 percent when the test is applied on two normal skewed distributions with unequal shapes. In all other cases, the difference in power between both tests is smaller (See table A1.1 to A1.9 in the additional file).

Considering the cases where sample sizes are unequal and SDR = 1, Student's-*t* test is sometimes better than Welch's *t*-test, and sometimes the reverse is true. The difference is small, except in three scenarios (See table A5.2, A5.5, and A5.6 in the additional file). However, because there is no correct test to perform that assures SDR = 1, and because variances are likely not to be equal in certain research areas, our recommendation is to always use Welch's *t*-test instead of Student's *t*-test.

To illustrate the differences in Type 1 error rates between Student's *t*-test and Welch's *t*-test, we simulated 1,000,000 studies under the null hypothesis (no difference between the means in each group) under four scenarios. We chose a small sample ratio (n_1 = 40 vs. n_2 = 60) to show that when the equal variances assumption was not met and SDR = 2, biased error rates are observed in Student's *t*-test. We compared Scenario 1, where the variance is the same in each group (SDR = 1; homoscedasticity assumption met) and sample sizes are unequal (See **v 2a**), with Scenario 2, where the variance differs between groups (SDR = 2) but sample sizes are equal ($n_1 = n_2 = 50$; see **Figure 2b**). Furthermore, we simulated Scenario 3, where both sample sizes and variances were unequal between groups and the larger variance is associated with the larger sample size (SDR = 2; see **Figure 2c**), and a similar Scenario 4, where the larger variance is associated with the smaller sample size (SDR = 0.5; see **Figure 2d**). *P*-value distributions for both Student's and Welch's *t*-tests were then plotted. When there is no true effect, *p*-values are distributed uniformly.

As long as the variances are equal between groups or sample sizes are equal, the distribution of Student's *p*-values is uniform, as expected (see **Figures 2a** and **2b**),

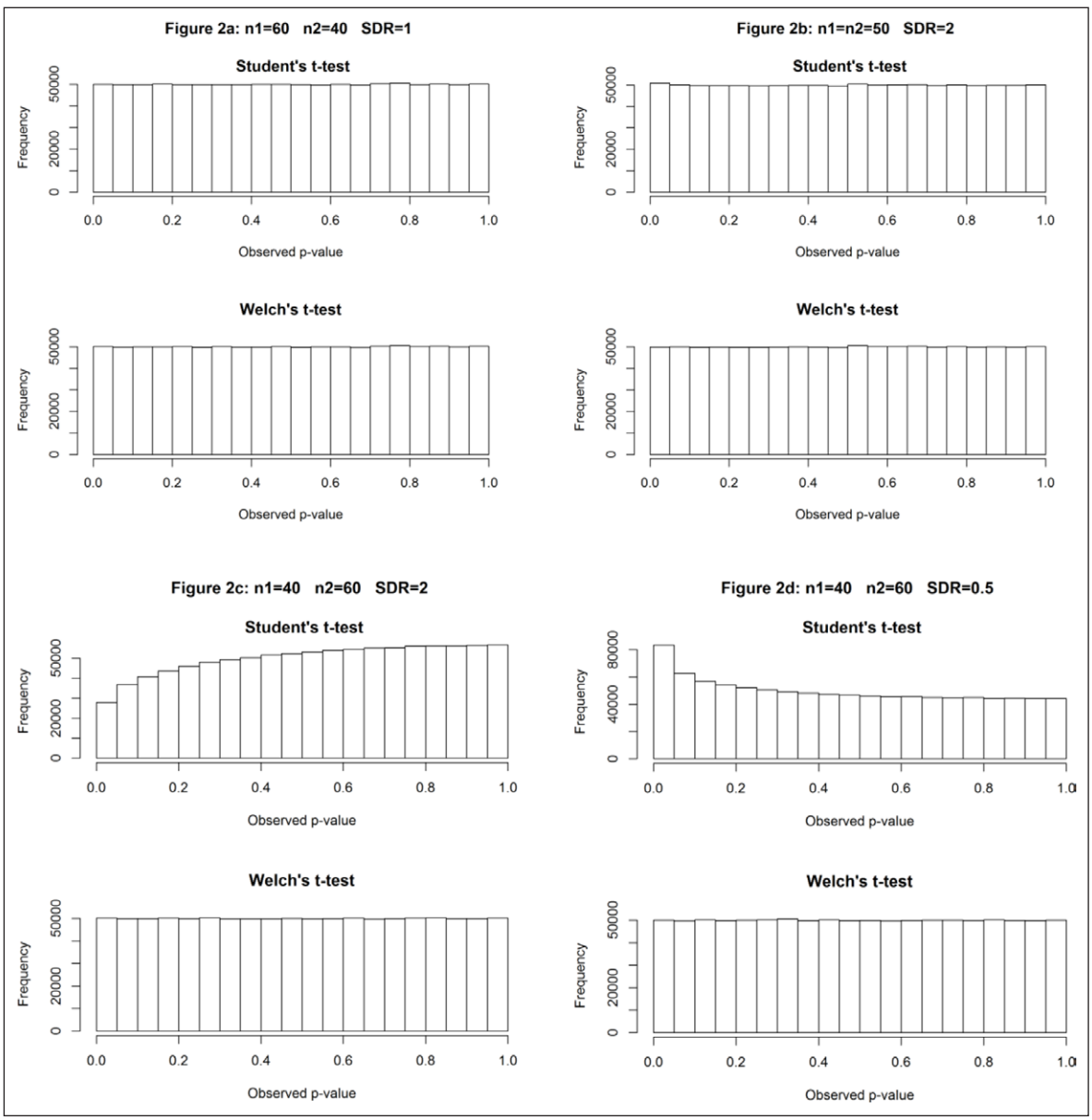

Figure 2: *P*-value distributions for Student's and Welch's *t*-test under the null as a function of SDR, and sample size.