1.

1. Yes, there is evidence of imbalance. The logit model contains significant coefficients for two consumption months and 420\_3, 43111\_3, 4701.2\_1. In the Quick Means test, all of the consumption variables had a t-statistic much larger than two, indicating that pre-trial consumption predicts placement into the treatment or control group. There were also several other variables that had t-statistics greater than two in the Quick Means test, which is further evidence that there is imbalance.
2. There are several differences between the logit results and the quick means test results. As stated above, only two consumption months were significant in the logit, while all were significant in the Quick Means Test. There were also many more significant variables in the quick means test than the logit regression. Because more variables are significant in the quick means test, this is likely due to multicollinearity. There are a lot of variables in the model, which increases the chance of a multicollinearity problem. In addition, the standard errors of these variables seem large relative to the coefficient. For several questions, some responses were significant in the quick means test while others were significant in the logit model (e.g. Questions 420 and 4701). This could be partly due to multicollinearity or perfect separation, but we can’t explain why some variables are significant in the logit model but not in the quick means test. Because many responses to the same question are significant in one of the two tests, these questions may require further investigation into their significance.
3. Even though the logit model can provide significance levels, there are many reasons estimates from a logit model can be misinterpreted that will not affect the results of the quick means test. For example, as the number of dummy variables increases, the probability of encountering multicollinearity problems also increases, which blows up the standard errors. This increases the chances of the logit model showing coefficients as insignificant when they are actually significant.
4. We think that the questions that gauge “enthusiasm” or questions that are too technical may be irrelevant or unreliable indicators. The program was opt-in, so all participants have some degree of interest. Participants may not know how to answer more technical questions like distinguishing between an immersion and instantaneous electric water heater in question 4701 (this question was significant in the logit model (p-value = 0.008) but insignificant in the quick means test (t-stat = 0.62); we’re not sure why this would happen).

7.

1. The coefficient of the treatment-trial interaction variable in the unweighted model is -0.0080. In the weighted model, the coefficient remained positive but increased in absolute value to -0.0253.
2. In the first regression without weights, the coefficient says that in the trial period, consumption in the treatment group was 0.80% smaller compared to the control group. However, this coefficient was not significant so there is no evidence that the treatment was effective at all.
3. In the second regression with weights, the coefficient says that in the trial period, consumption in the treatment group was 2.53% smaller compared to the control group. This coefficient was significant at the 1% level, showing the treatment was effective.
4. Because we found in Section 1 that the data is biased, the weighted regression coefficient is more valid than the unweighted regression coefficient. If the data was not biased, the weighted and unweighted regressions would give the same result. Since the results were different, the weighted regression coefficient is believable.