Supplementary materials - methods: Sample Size and Power Calculation

We performed a simulation study to determine the power of various study designs to detect a shift in the time (number of weekly challenges) to infection between the treatment and control groups of macaques given weekly challenges using the log-rank test. A key question we considered is whether increasing the number of macaques in each group meaningfully improves statistical power. Because we do not have a specific *a priori* value for the true shift, we considered a variety of scenarios.

In each scenario, we assumed the number of macaques in control (n_1) and the treatment groups (n_2) to be equal. We set the mean (number of weekly challenges) to infection as μ_1 = 9 or 10 in the control group and μ_2 = 12 or 13 in the treatment group, giving a shift (difference) of δ = 2, 3, 4, 8, and 9 challenges to represent potential real scenarios. We set the common standard deviation in the number of weekly challenges to infection to $\sigma_1 = \sigma_2$ = 1, 2, 4, and 5. In each simulated trial, we generated random normal data from the above means and standard deviations and rounded each number up to the nearest positive integer, corresponding to the way data are observed in whole weeks. We simulated 10,000 data sets for each scenario, computed log-rank test p-value for each, then compared survival times of the treatment and control groups. Next, we computed the average rejection rate at the 0.05 significance level for each scenario to obtain the simulated power. Supplemental table 1 displays the sample size and power calculations for the different combinations of the mean number of weekly challenges for each group (μ_i), the difference between the two group means (δ), the standard deviation of challenges within each group (σ_i), and the sample size of each group (n_i). In most of our recent studies carried out in our lab, we observed that the shift between the treatment and control groups were much larger than 4 but standard deviations for both groups were small. This indicates that we had >80% power for most scenarios even when we had 6 macaques in both treatment and control groups. For some extreme scenarios, we might have less than 80% power with small sample size.