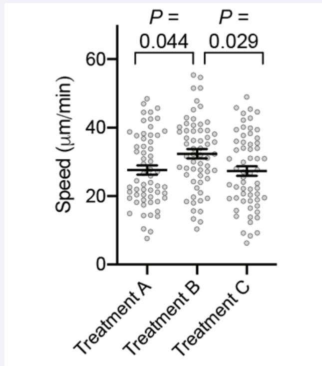
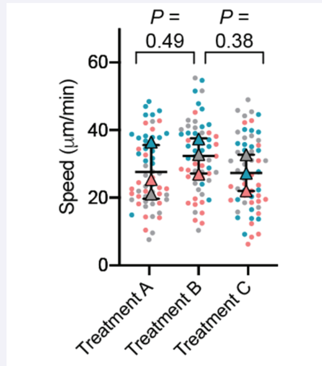
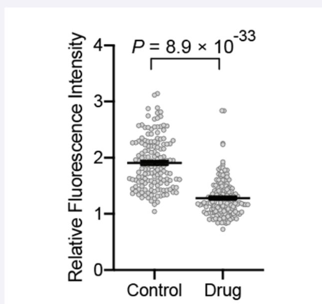
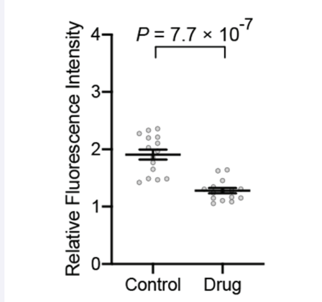
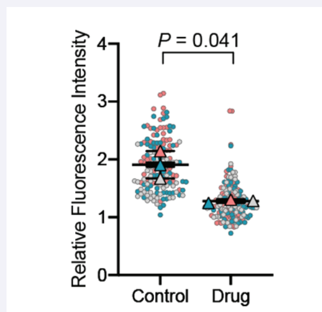
Experiment	If *n* = number of observations	If *n* = number of experiments	Potential outcomes and problems
Researchers compare nucleus size within tumors vs the surrounding cells. They look at 100 transformed and normal cells in 5 different tissue samples.	*n* = 500 They plot the nucleus size for each cell as a beeswarm plot with tiny error bars and P values.	*n* = 5 They compare the nucleus size in transformed and normal cells, paired by tissue sample. Now they have five biological replicates, each encompassing 100 technical replicates. They perform a paired test and get a reasonable P value.	By using an *n* of 500, it is likely that the P value will be artificially smaller. By averaging the technical replicates and using an *n* of 5, the researchers can confirm if the observed difference across patient samples is larger than the natural variation within any single tissue sample.
In an in vitro experiment, researchers compare the rates of filament growth of actin orthologs.	*n* = 1000 They measure the growth rate of thousands of filaments from each organism. They then calculate the statistics using the number of filaments as *n*, resulting in a tiny P value.	*n* = 4 They repeat the experiment over multiple days, with different stocks of purified proteins. They calculate the mean filament growth rate in each run and report a P value based on those means.	Counting each filament as a separate sample provides an estimate of the inherent variability of filament growth rate within that one sample. But using that to then compare two orthologs on different coverslips is not appropriate, because there is nothing controlling for different handling of the two proteins during imaging. Repeating the experiment with different stocks of protein takes into account random and systematic errors including concentration/pipetting error, room temperature fluctuations, protein degradation, and other, less foreseeable, variables. Readers would want to know if any findings reported can be replicated under similar conditions, for example in another lab.
Researchers use laser ablation to cut individual spindle fibers, then observe the downstream effects in those same individual cells. In other cells in the same sample, they apply the same laser dose, but not directed at spindle fibers.	*n* = 30 They perform the treatment or control on 30 different cells, approaching each cell as an *n* and calculating a P value to compare the two treatments.	*n* = 4 They repeat the experiment over multiple days, with different stocks of purified proteins. They calculate the mean filament growth rate in each run and report a P value based on those means.	Because each cell can be randomly assigned to the control or treated group, it is appropriate in this case that cell is its own sample and calculate a P value from one day's experiment using *n* as the number of cells. The researchers may conclude that cells treated with laser ablation differ significantly from those that don't have their spindle fibers cut. But that may only be true for that cell strain at that particular passage number and at a specific temperature of the scope room. Observing similar results over multiple days or even with different cell types makes this claim robust.

Experiment	If *n* = number of observations	If *n* = number of experiments	Potential outcomes and problems
A researcher measures the speeds of 20 crawling cells per condition, and repeats the experiment on 3 different days.	*n* = 60 cells Test: ANOVA+ Tukey post test 	*n* = 3 experiments Test: ANOVA+ Tukey post test 	Here, the researcher would come to a different conclusion based on what they consider *n*. In this example, data were collected on cells prior to treatment, so they are all untreated. Therefore, differences are due to unpredictable sample-to-sample fluctuations. This example highlights that these chance differences are amplified when each cell is considered its own experiment. When comparing actual treatments, using *n* = 60 could lead to erroneously small P values.
A researcher measures the pixel intensity of actin staining in cell protrusions. The researcher measures 10 protrusions per cell for 5 cells, and repeats the experiment 3 times.	*n* = 150 protrusions (top), or 15 cells (bottom) Test: unpaired t-test  	*n* = 3 experiments Test: paired t-test 	In this example, there are three choices for *n*: the number of protrusions, the number of cells, or the number of experiments. In this case (Velle and Campellone, 2018), the P value is less than 0.05 regardless of which *n* is used for statistical analysis. This example illustrates that while a miniscule P value could raise red flags, the underlying conclusions may still be appropriate.

Velle, K.B., and K.G. Campellone. 2018. Enteropathogenic E. coli relies on collaboration between the formin mDia1 and the Arp2/3 complex for actin pedestal biogenesis and maintenance. *PLoS Pathog.* 14:e1007485. doi:10.1371/journal.ppat.1007485.