A GUIDE FOR RUNNING STATISTICALLY POWERED CLONE-SCREENING EXPERIMENTS

Kris Tyner, Sneha Jani, Cayley Hoyer, Chelsea Edgar and Katie Madler
Culture Biosciences, 269 E Grand Avenue, South San Francisco, CA 94080, USA
www.culturebiosciences.com

Summary

The primary goal of clone screening in bioreactors is to identify clones with improvements in key performance indicators (KPIs) such as titer, rate and yield over the course of a bioprocess. Since all bioprocesses have an inherent level of variability, replicates of each clone should be run in every experiment. Limited bioreactor capacity is often a primary driver of how replicates are structured rather than the desired replicate count driving bioreactor usage. In this white paper, we describe how conducting a power analysis can determine the correct sample size for a statistically powered clone-screening experiment.

INTRODUCTION TO POWER ANALYSIS

The number of replicates required for any statistically powered experiment is dependent on the variability of the data, the magnitude of difference that is to be detected and the tolerance for false-positive and false-negative results in the data. Since these parameters can vary widely between different programs and over the course of a cell line development program, the necessary sample size should be assessed on a regular basis.

The required number of replicates can be calculated mathematically using a **power analysis.** A power analysis considers bioprocess variability, the magnitude of the desired effect and the false-positive and false-negative result tolerance when determining the number of replicates required for an experiment. Each of these factors is described in more detail below.

(A) Variability

The first factor to consider when determining the number of replicates required for an experiment is the variability of the bioprocess. If the results of replicate experiments from a bioprocess are normally distributed, there will be a range of values distributed around the mean. For example, if a cell line results in a final titer of 100 g/L and a standard deviation of 3 g/L (illustrated as Process A of Figure 1), the results from 10 replicates would be expected to fall within a narrower range of absolute titers than if the process had a standard deviation of 5 g/L (Process B of Figure 1). While the true performance of the cell line is the same, the higher variability in Process B leads to a wider distribution of absolute titers in replicate experiments.

Figure 1: *Hypothetical titers obtained in a bioprocess that has a standard deviation of 3 g/L (A) compared with a process with a standard deviation of 5 g/L (B).*

(B) Effect size

The next factor to consider when determining replicates is the magnitude of performance change that is required from cell line improvements, also known as the effect size. When scientists are interested in a large effect size, they can have higher confidence in a positive result and therefore require fewer experimental replicates.

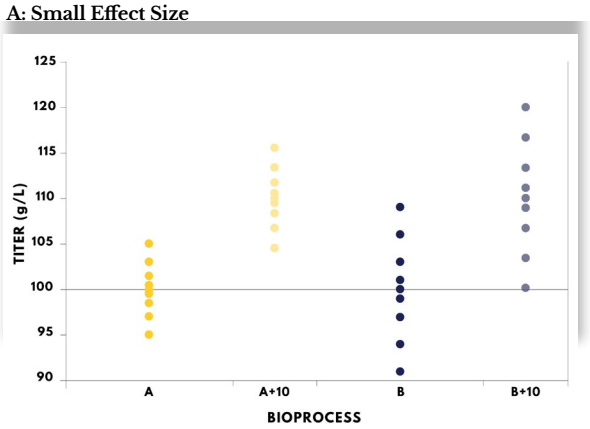
Early cell line modifications often yield large relative improvements in cell line performance as genetic edits are frequently rational, on-pathway edits (a.k.a. the low-hanging fruit). As such, these experiments typically have a higher effect size. On the other hand, as overall cell line performance improves over time due to the combination of desirable genetic edits, the relative improvement attributed to each new genetic edit may begin to diminish over time. For these experiments, a smaller effect size is typical, and to compensate for a lower confidence, replicates may be increased accordingly.

It is important to note that some effect sizes may not be practical when considering the goals of a program. For example, if the initial titer in a bioprocess is 15 g/L and the program goal is 100 g/L, an improvement of 0.5 g/L in magnitude may not be a practical effect size to detect due to the number of replicates required to detect the change and the number of sequential cell line consolidations that would be required to reach the program goal. Careful consideration of the effect size is important to balance between detecting improvements with fewer replicates and failing to identify improvements below the effect size threshold.

The variability of a process impacts the ability to detect improvements, and this is especially relevant when the magnitude of the change is small. When the variability of a process is low, it is easier to detect smaller improvements in cell line performance, and fewer replicate runs are required in order to achieve statistical significance. Returning to the hypothetical bioprocesses A and B from Figure 1, an example of cell line performance detection is illustrated in Figure 2A. If an improved cell line produces 110 g/L of product compared to the 100 g/L produced by the control line, the variability in the process will impact the ability to detect the improved cell line. Here, you can see that with 10 replicates, it is relatively easy to visualize the difference between the 100 g/L cell line and an improved 110 g/L cell line when the variability is ±5% (process A vs. A + 10) as compared to a variability of ±10% (process B vs. B + 10). This increase in titer change is statistically significant with both levels of process variability but is clearer when the process variability is low (process A + 10 vs. B + 10).

Similarly, if the magnitude of the change that is expected is larger, the differences in variability between the processes become less pronounced. Figure 2B illustrates how an increase in titer from 100 g/L to 150 g/L is easy to see in either scenario (process A vs. A + 50 and B vs. B + 50). Here, the variability of the process is less relevant in the interpretation of the data because the magnitude of the change is so large. Fewer replicate experiments could be run to detect this difference.

A: Small Effect Size



B: Large Effect Size

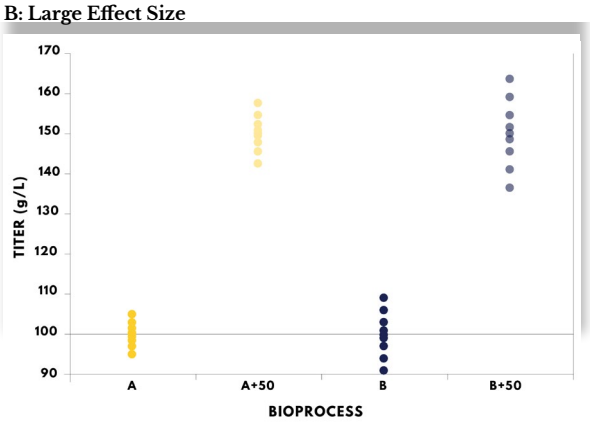


Figure 2: *Effect of variability on detecting a small (A) or large (B) effect size. The effect size in Panel A increases 10 g/L between bioprocesses A vs. A +10 and B vs. B + 10, respectively. The effect size in Panel B increases 50 g/L between bioprocesses A vs. A +50 and B vs. B + 50, respectively. Processes A, A + 10 and A + 50 have a variability of ±5% and processes B, B +10 and B + 50 have a variability of ±10%.*

(C) Tolerance of false-positive and false-negative results

Another factor that affects the number of experimental replicates is the tolerance of false-positive and false-negative results. While it is nearly impossible to eliminate false positives and false negatives from experimental datasets, replicates can be adjusted to balance the number of anticipated false-positive and false-negative results with the available throughput. This is an especially important consideration for bioprocess scientists who are faced with limited capacity to run large numbers of replicates, even when using a service such as Culture's.

A false-positive result occurs when a cell line appears improved in a test while its true performance is not actually improved (Figure 3). In a small number of replicates, it is possible for results to be on the higher end of the expected range purely by chance. However, if more replicates are included in the experiment, the average performance of the replicates may be more reflective of the true performance of the cell line.

A false positive is also known as a type I error. The tolerance of type I errors in a power analysis is represented by the Greek letter alpha (α). Alpha is a number between 0 and 1 that describes the probability of a type I error. Simply put, an alpha of 0.05 would mean that there is a 5% chance of a result being a false positive. The lower the value of alpha, the lower the chance of obtaining a false-positive result. It is important to remember, however, that lowering alpha requires increasing the required sample size.

		PERFORMANCE IN REALITY	
		IMPROVED	NOT IMPROVED
TEST RESULTS	IMPROVED	Test strain is improved over control strain i.e., True Positive	Test strain appears to be improved over control strain i.e., False Positive (Type I error, **α**)
	NOT IMPROVED	Test strain appears to not be improved over control strain i.e., False Negative (Type II error, **β**)	Test strain is not improved over control strain i.e., True Negative

Figure 3: *Distinguishing between true-positive, false-positive, true-negative and false-negative results based on experimental (test) results and performance in reality.*

Conversely, a false-negative result is one where a cell line demonstrates no improvement in performance when its true performance is, in fact, improved over control (Figure 3). This can occur when the observed performance of a cell line is on the lower end of the expected range. Again, if more replicates are included in an experiment, the average performance of the replicates may be more reflective of the true performance of the cell line. The probability of obtaining a false-negative result is represented by the Greek letter beta (β). The power of an assay, or the probability of obtaining a true positive, is calculated as 1 - **β**.

Together, assay variability, effect size and tolerance of false-positive and false-negative results can be used in a power analysis to determine the number of replicates required in an experiment. The required sample size for a statistically powered study is calculated by the following equation:

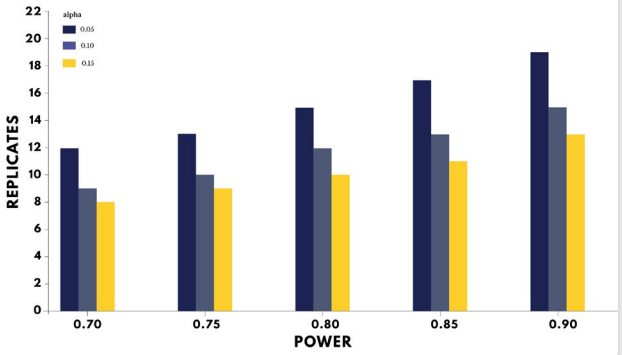
$$n = \frac{(Z_\alpha + Z_\beta)^2 \times \sigma^2}{d^2}$$

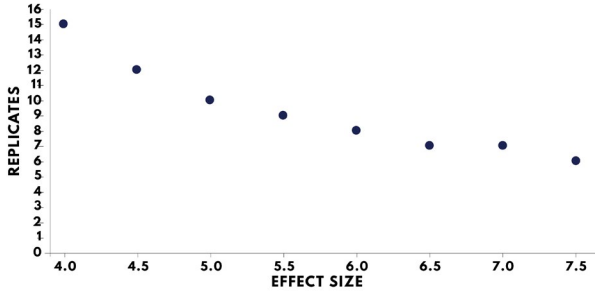
Where:

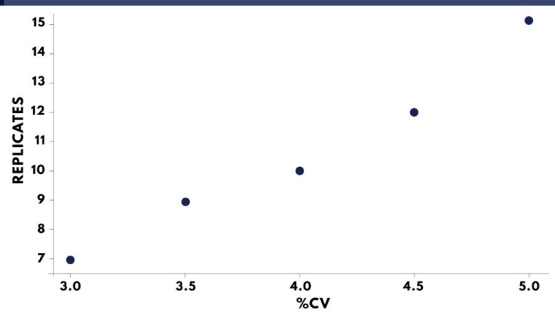
n= sample size
d= effect size
α = significance level of the test
β = 1 - power
σ = standard deviation of the process
Z= standard score on a normal distribution (how many standard deviations from the mean each data point lies)

PERFORMING A POWER ANALYSIS

As an example, a control bioprocess with a 100 g/L titer and a 5 g/L standard deviation in titers is used for screening cell lines, and the variation is assumed to be equivalent across cell lines. The desired increase in titer to be detected is 4 g/L. Using an alpha of 0.05 (the probability of obtaining a false-positive result) and a power of 0.8 (the probability of observing an effect), a one-tailed power analysis (one-tailed because the cell line is predicted to improve performance) determined that 10 replicates would be required to detect a performance change of this magnitude given these variables.

CHOICE	EXAMPLE
Adjust the tolerance of false-positive (alpha) and false-negative (beta) results By adjusting alpha and beta, the number of replicates required can be decreased. The graph in the right panel assumes an effect size of 4 and standard deviation of 5.	

CHOICE	EXAMPLE
Increase the effect size As shown in the graph to the right, increasing the magnitude of the improvement to be detected from 4.0 to 7.5 lowers the number of replicates required to detect cell line effect. This example assumes standard alpha (0.05) and beta (0.8) values.	

CHOICE	EXAMPLE
Decrease the variability of the bioprocess Implementing improvements that target sources of process variability can decrease the number of replicates required. The example in the right panel illustrates how reducing process variability to as low as 3% coefficient of variation (CV) can decrease the number of replicates required in a statistically powered study.	

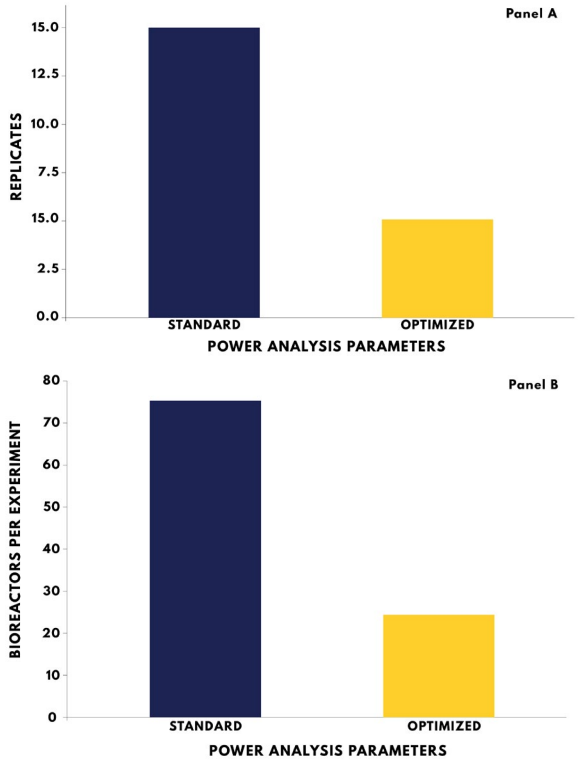
CHOICE	EXAMPLE
A combination of approaches Several bioprocess factors can be adjusted in parallel to balance the number of replicates required with the available throughput of bioreactors. Panel A illustrates how adjusting the alpha to 0.1, beta to 0.75, and effect size to 5 and decreasing the variability of an example process to a coefficient of variation (CV) of 3.5% reduced the required number of replicates required from 15 to 5. To illustrate the time and cost savings a three-fold reduction in sample size can have in cell line-screening programs, Panel B depicts how screening 5 cell lines plus a control using the adjusted bioprocess of Panel A would reduce the number of required bioreactors from 75 to 24.	

Table 1: *Factors that can be adjusted to decrease the number of replicates required in a statistically powered study. The left side of the table describes how to decrease sample size without severely compromising the power of a study by using hypothetical examples. The right side of the table graphically illustrates the effects of each adjustment on sample size.*

EFFECTIVE CELL LINE SCREENING

The constraints of a strain screening program, combined with the expected effect size, desired statistical power and observed variation of a bioprocess, often require a sample size exceeding available bioreactor capacity. While optimizing the inputs to a power analysis can reduce the number of replicates required in an experiment, bioreactor capacity is often a limiting factor.

Culture can provide the bioreactor capacity required to screen for desired cell line improvements with sufficient statistical power. Importantly, statistically powered experiments will ensure that improved strains are reliably detected while minimizing false-positive results. Studies designed with a robust statistical strategy are also more efficient, as cell lines with improved performance are reliably detected and resources are not spent further developing false-positive cell lines. Cell line improvement programs, often run with tight timelines and ambitious goals, simply cannot afford to run statistically underpowered screening studies.

Are you running enough replicates in your bioprocess to make statistically significant decisions? Check out **Culture's high-throughput bioreactor lab services** to learn how you can run statistically powered experiments that do not need a tradeoff between structuring replicates based on statistics and throughput of cell lines.

Culture is your upstream bioprocess lab, in the cloud.

CONTACT US
inquiries@culturebiosciences.com
culturebiosciences.com