Power-based sample size calculation is a crucial part of study design that used hypothesis testing for inference and should be performed at the time of planning a study. It is advised to contact your statistician as early as possible. Too small a sample may not be sufficiently powered to detect a difference between groups. Available subjects, cost, and time also need to be considered but it is not worth conducting an underpowered study. Too large a sample may waste time and money and could expose more participants to unnecessary risk.

There are several parameters required for a power-based sample size calculation:

1. **Type I error (α)**, which is the probability of falsely rejecting the true null hypothesis and is typically 0.05 or 0.01. The larger the alpha the smaller the necessary sample size.
2. **Power (1-β)**, which is the probability of correctly rejecting the true null hypothesis and is typically 0.8 or 0.9. The smaller the power the smaller the necessary sample size.
3. **Minimal detectable difference**, which is the smallest difference in the outcome between the study groups (effect size) that is of scientific interest and is typically specified by the investigators. The larger the effect size the smaller the necessary sample size.
4. **Variance**, which is the variability in the outcome and is typically estimated from a previous study or pilot data. The smaller the variability the smaller the necessary sample size.
1) Example scientific question for comparing two independent proportions:

We are planning a study to compare two different drugs, A and B, and hypothesize that drug A will be better than drug B, in terms of higher HIV-1 viral load suppression. We plan to identify a random sample of HIV-1 infected subjects and randomly assign them to one of the two drugs. At the end of six months we will measure the subject’s HIV-1 viral load and compare the two drugs. We expect to see a 20% difference in the proportion HIV-1 viral load suppression between the two drugs. Assuming equal sized groups, an $\alpha = 0.05$ and Power = 0.80 the below output suggests we need to study 62 subjects per drug to detect a 20% difference in HIV-1 viral load suppression. The proportion of HIV-1 viral load suppression will be compared between drugs using a Pearson’s chi-square test.

```
. power twoproportions 0.7 0.9, p(0.8) a(0.05)
```

Performing iteration ...

Estimated sample sizes for a two-sample proportions test
Pearson's chi-squared test
Ho: $p_2 = p_1$ versus Ha: $p_2 \neq p_1$

Study parameters:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>alpha</td>
<td>0.0500</td>
</tr>
<tr>
<td>power</td>
<td>0.8000</td>
</tr>
<tr>
<td>delta</td>
<td>0.2000 (difference)</td>
</tr>
<tr>
<td>$p_1$</td>
<td>0.7000</td>
</tr>
<tr>
<td>$p_2$</td>
<td>0.9000</td>
</tr>
</tbody>
</table>

Estimated sample sizes:

<table>
<thead>
<tr>
<th>N</th>
<th>124</th>
</tr>
</thead>
<tbody>
<tr>
<td>N per group</td>
<td>62</td>
</tr>
</tbody>
</table>
2) Example scientific question for comparing two independent means:

We are planning a study to compare two different drugs, A and B, and hypothesize that drug A will be better than drug B, in terms of lower HIV-1 viral load. We plan to identify a random sample of HIV-1 infected subjects and randomly assign them to one of the two drugs. At the end of six months we will measure the subject’s HIV-1 viral load and compare the two drugs. We expect to see a 0.5 log difference in the average HIV-1 viral load between the two drugs. We also assume that the standard deviation of HIV-1 viral load to be 0.5 logs for both drugs. Assuming equal sized groups, an $\alpha = 0.05$ and Power = 0.80 the output below suggests we need to study 17 subjects per drug to detect a 0.5 log difference in average HIV-1 viral load. The average HIV-1 viral load will be compared between drugs using a t-test.

```
. power twomeans 3 2.5, p(0.8) a(0.05) sd(0.5)
Performing iteration ...
```

Estimated sample sizes for a two-sample means test

```
t test assuming sd1 = sd2 = sd
Ho: m2 = m1 versus Ha: m2 != m1
```

Study parameters:

```
alpha = 0.0500
power = 0.8000
delta = -0.5000
m1 = 3.0000
m2 = 2.5000
sd = 0.5000
```

Estimated sample sizes:

```
N = 34
N per group = 17
```

Sample size calculations become more complicated when study participants are not independent. Sample size calculations for a cluster randomized trial like a Stepped Wedge design are very complex and a statistician should be consulted.

Please contact the CFAR Biometrics Core for specific questions or issues you have about your power calculations.