Suppose you have N subjects who each run M trials. The score for the ith trial for the jth subject is defined by

$$X_{ij} = \mu + \alpha_j + \epsilon_{ij} \tag{1}$$

where μ is the overall grand population mean across subjects, α_j is the deviation from the grand mean for subject j, and ϵ_{ij} is noise for a particular trial for a particular subject.

μ is an unknown fixed value. α_j varies across subjects that might be selected for the experiment. Assume the subject population is distributed normally with $\mathcal{N}(0, \sigma_\alpha)$. The ϵ_{ij} term describes variability within a subject across trials. We assume that it also follows a normal distribution, $\mathcal{N}(0, \sigma_\epsilon)$, which has a different standard deviation than for variability across subjects.

Equation (1) is sometimes called a *random effects model*. Oftentimes data from a random effects model are tested with a *fixed effect model*. A common approach is as follows:

The mean of a subject's trials becomes an individual score, Y_j. The set of N Y_j values are then fed into a one-sample t test ($df = N - 1$). The variability across subjects will be

$$\sigma^2 = \sigma_\alpha^2 + \frac{\sigma_\epsilon^2}{M} \tag{2}$$

because M trials contribute to the subject's average score. Larger σ_ϵ leads to more variability and makes it more difficult to reject the null, but the contribution is divided by M, so it may not have a big effect. When computing the standard error for the t-test, one uses the number of subject scores, so

$$\sigma_{\overline{X}} = \sqrt{\frac{\sigma^2}{N}} = \sqrt{\frac{\sigma_\alpha^2}{N} + \frac{\sigma_\epsilon^2}{MN}} \tag{3}$$

Notice that the σ_ϵ term is divided by MN, so it contributes less variability than a corresponding σ_α term. In a fixed effect model analysis, any observed variability is assumed to come from σ_α, which is not necessarily correct but can often be close to accurate.

Simulated experiments with a fixed number of trials

Suppose an experiment is being run that can gather $M \times N$=100 trials. How should the experimenter distribute those trials across N subjects? There may be practical issues that limit how many trials each subject can produce and how many subjects are available; but from a statistical perspective, it is almost always best to maximise the number of subjects (e.g, $N = 100$ and $M = 1$).

I ran simulated experiments that set (N, M) to be $(100, 1)$, $(50, 2)$, $(20, 5)$, $(10, 10)$, or $(2, 50)$. For each subject, the simulation generated a value α_j from a normal distribution with mean $\mu = 0.5$ and $\sigma_\alpha = 1$. For each subject, I generated M trials by drawing random values of ϵ_{ij} from a normal distribution with a mean of zero and a standard deviation of 0.25, 0.5, 1, 2, or 4. These values define individual trial scores using equation (1), above.

The subject scores were tested using a fixed effect model (a standard one-sample t-test for $H_0 : \mu = 0$). Table 1 shows the proportion of 10,000 simulated experiments that rejected the null (estimated power). There are two notable effects. First, as σ_ϵ increases, power decreases.

σ_ϵ	N	M	Estimated Power
0.25	100	1	1
0.25	50	2	0.93
0.25	20	5	0.57
0.25	10	10	0.29
0.25	2	50	0.06
0.5	100	1	0.99
0.5	50	2	0.91
0.5	20	5	0.56
0.5	10	10	0.29
0.5	2	50	0.06
1	100	1	0.94
1	50	2	0.81
1	20	5	0.49
1	10	10	0.27
1	2	50	0.06
2	100	1	0.61
2	50	2	0.52
2	20	5	0.36
2	10	10	0.23
2	2	50	0.06
4	100	1	0.23
4	50	2	0.21
4	20	5	0.18
4	10	10	0.15
4	2	50	0.06

Table 1: Results of simulated experiments with a fixed number (100) trials distributed across a N subjects for varying σ_ϵ.

This makes sense, because increased variability across trials means larger standard error in equation (3) and thus smaller t statistics and smaller effect sizes. Second, Table 1 shows that for a fixed value of σ_ϵ, the maximum power is always for $N = 100$, $M = 1$, which will minimise the standard error in equation (3). This makes sense because in equation (3) the σ_ϵ term is divided by MN, which always equals 100 trials for these simulated experiments. Thus, to minimise standard error, a researcher should increase N at the expense of decreasing M. An additional advantage of increasing N is that the t-test uses a degrees of freedom ($df = N - 1$), so a larger N reduces the t critical value.

Simulated experiments with a fixed number of subjects that vary in number of trials

The above analysis and simulations demonstrate that one cannot just trade off subjects for trials. Everything else equal, it is always best to maximise subjects. However, subjects are

σ_ϵ	M	Estimated Power
0.25	1	0.73
0.25	20	0.75
0.25	50	0.75
0.25	100	0.75
0.25	1000	0.75
0.5	1	0.66
0.5	20	0.75
0.5	50	0.76
0.5	100	0.75
0.5	1000	0.75
1	1	0.47
1	20	0.73
1	50	0.75
1	100	0.75
1	1000	0.76
2	1	0.22
2	20	0.68
2	50	0.72
2	100	0.74
2	1000	0.76
4	1	0.1
4	20	0.52
4	50	0.63
4	100	0.69
4	1000	0.76

Table 2: Results of simulated experiments with a fixed number of subjects (N=30) with a varying number of trials (M) and σ_ϵ.

often more difficult to acquire than trials, and for a fixed number of subjects there is an advantage to gathering more trials. However, the advantage is sometimes modest. Table 2 shows the proportion of 10,000 simulated experiments that reject the null. Each experiment used $N = 30$ and varied the value of M and the value of σ_ϵ. In general larger M values gives rise to higher estimated power (this does not always happen due to random sampling in the simulations). The advantage is largest when σ_ϵ is big, which makes sense given the properties of equation (3).

Note that even with $\sigma_\epsilon = 2$, which is twice the size of σ_α, an increase from $M = 50$ to $M = 1000$ only leads to a power increase of 0.02. So increasing the number of trials by a factor of 20 hardly benefits the researcher but greatly bothers the subjects. There are situations where increasing M has large effects. For $\sigma_\epsilon = 1$, increasing M from 1 to 20 produces a 0.26 increase in power.

Unless σ_ϵ is much larger than σ_α, there is little gain to gathering more than 20 or 50 trials per subject.