My model for scientific investigations: Question $\underset{\text{Design}}{\longrightarrow}$ Data $\underset{\text{Analysis}}{\longrightarrow}$ Conclusion

Why design matters:

1. randomized experiment or observational study ⇒type of conclusion
   On average randomized treatments ⇒only diff. is the treatment
   if find a difference ⇒causal claim (treatment caused the difference)
2. design ⇒ appropriate analysis
3. bad design sinks a study
   fate of 187 ecology manuscripts in Scandinavia.
   27 rejected. Majority because of bad design

How many replicates? Motivating study:
Walk-With-Ease (WWE), national program to increase elderly physical activity (PA)
Does having an individual health coach provide a benefit?
individuals randomly assigned to receive a health coach or not
response is minutes of moderate-vigorous PA in a week
How many individuals?

5 ways to answer

1. tradition, usually n=3 per treatment
2. as many as you can afford
3. precision of an estimate
4. width of a confidence interval
5. power of a hypothesis test

Standard deviations and standard errors (reminder)
sd: variability between observations. Describes the data
se: precision of an estimate. Describes the estimate

$$\text{se mean} = \frac{s}{\sqrt{n}} = s\sqrt{\frac{1}{n}}$$

$$\text{se difference} = s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Statistical criteria require information from the client
se, ci width: need the sd and desired result
power: need the sd and a difference that matters

Sample size using precision:
pilot data gives sd = 14.8 minutes/day
want se of a mean = 2 minutes/day
result is n=54.8, round up 55, per group
Studies on people: increase to account for dropout

e.g. expect 20% dropout, n = 69
want se of a difference = 2 minutes per day
assume $n_1 = n_2$
result is n = 109.5 (round to 110) per treatment
check: se = $s * \sqrt{(}2/110) = 1.99$

Sample size using confidence interval (ci) width:
For an estimate that is normally distributed with an se that depends on an estimated sd

$$\left[\hat{\theta} - (t_{1-\alpha/2,\, df})(\text{se }\hat{\theta}),\;\; \hat{\theta} + (t_{1-\alpha/2,\, df})(\text{se}\hat{\theta})\right]$$

95% ci: α = 1-coverage = 0.05, want 0.975 quantile of the t distribution
df is the amount of information in s
one mean: $n - 1$
difference of two means: $n_1 + n_2 - 2$
goal: a confidence interval width of 4 for mean, e.g. a ci of (18, 22) or (-1, 3)
No closed form equation for n, try various values of n and close in on the solution
result is 213 per group
Why so much larger?
ci width is $2 * (t_{1-\alpha/2,\, df})$ times the se, approx 4 se

Power and review of hypothesis tests:
Type I error: reject H0 | H0 true = bad test result. Has probability α
Stat theory designs a test to have a specific α
Related to p-value. α specified in advance, p computed from the data
p = P[as or more extreme result by chance when H0 true]
Type II error: accept H0 | H0 false = bad test test. Has probability β
Two decisions (test accepts or rejects) and two states of nature (H0 or Ha)

	Test result, based on data	
Truth	accept H0	reject H0
H0 true	good	type I error, α
H0 false	type II error, β	good, power

Power = $1 - \beta$ = P[test rejects H0 | there is a non-zero difference]

Sample size using power:
5 related quantities: α, power = $1 - \beta$, difference = δ, n, and design
fundamental equation:

$$\delta = \left(t_{1-\alpha/2,\, df} + t_{power,\, df}\right) se = \left(t_{1-\alpha/2,\, df} + t_{1-\beta,\, df}\right) se$$

note: uses a "shifted" t approximation to a non-central t distribution
computer software uses the theoretically correct non-central t
the shifted t approx. is really good, except for very small n, e.g. $n < 5$
Derivation and explanation: see powerpictures.pdf notes