Sampling & Power Calculations

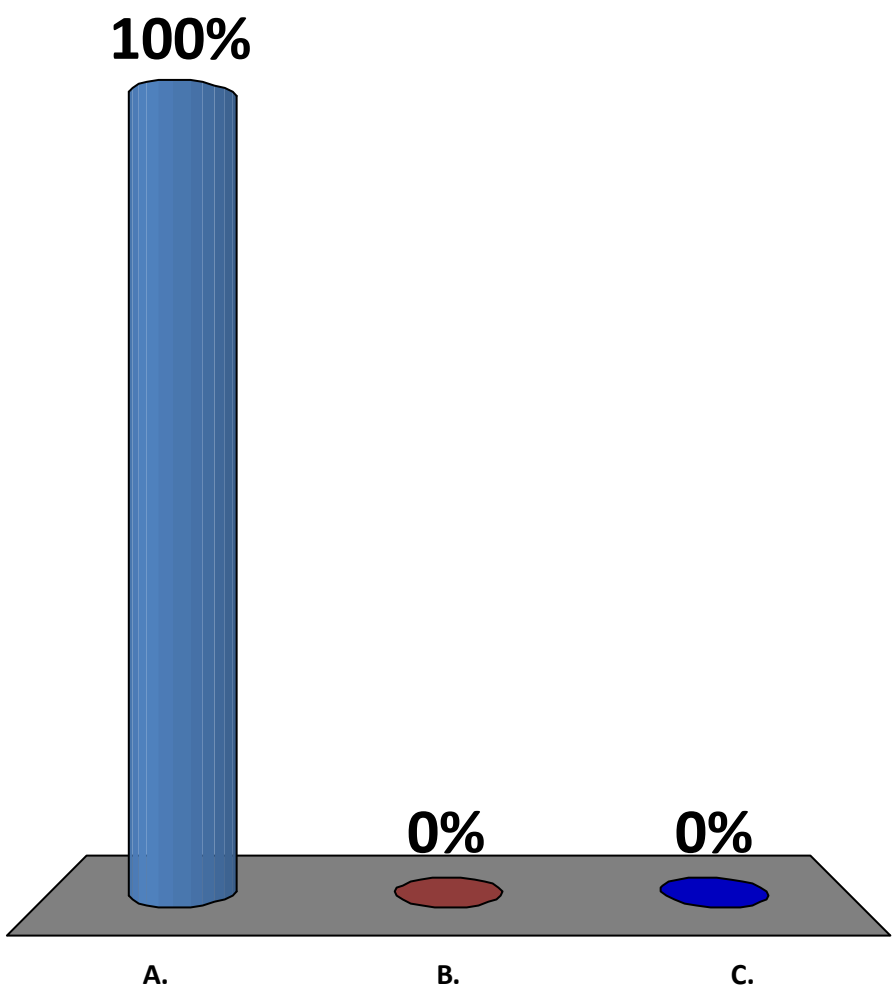
Maria Jones

18 May 2017

ID4D IDENTIFICATION FOR DEVELOPMENT
i2i DIME TRANSFORM DEVELOPMENT
BILL & MELINDA GATES foundation
THE WORLD BANK
UKaid from the British people

Are you ready to learn about sampling?

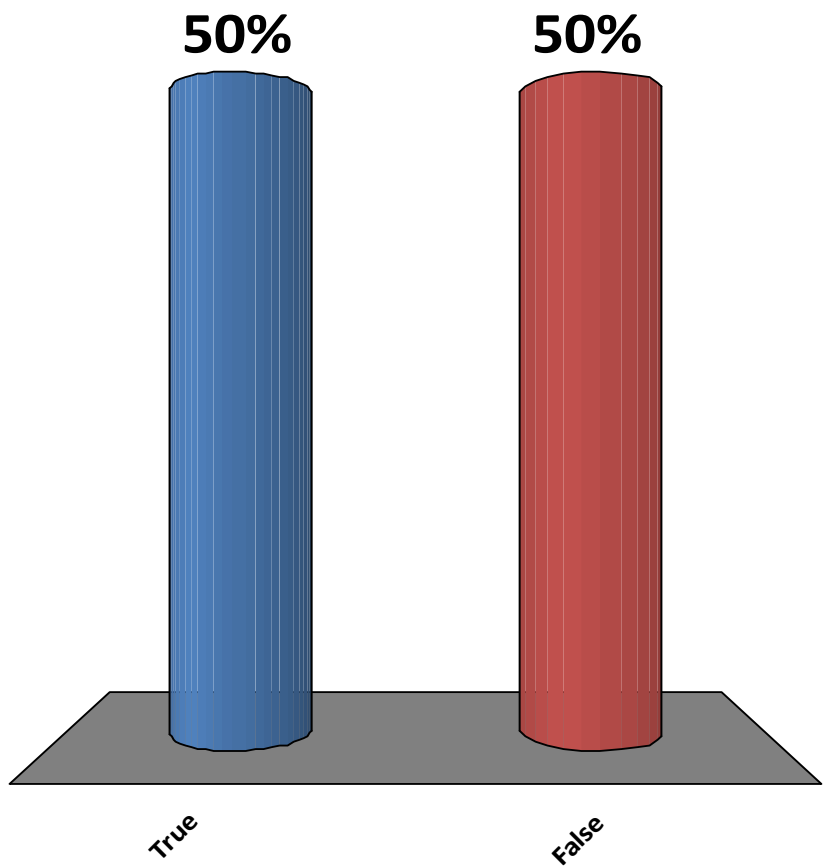
A. Yes, sampling is fun!
B. No, no, no
C. What? I'm not awake yet



Randomized assignment of an intervention is the same as random sampling

A. True

✓ B. False



Introduction

- Think of sample size as the **accuracy of a measuring device**

- The more observations you have
  - The more precise is your "measuring device"
  - The more confident you are about your conclusions

- The more complicated the question you want to answer, the more data points you need

Introduction

- Imagine you had to sample letters to “estimate” what the below sentence says

- # of revealed letters is like the # of observations
  - Say each letter costs US$ 100,000
  - You don’t want to spend all your budget on letters but you don’t want to guess wrong!

I			i		e		I					
	E							i				

Introduction

With a larger sample, you can be more confident you make the right inference:

I		l	i		e		I	m	p		c	t
	E	v		l			t	i	o	n		

Introduction

- The more observations the better, but we all have budget constraints
- How to determine how many letters is ‘enough’? What sample is sufficient for your research question?
- From today’s session, learn how to start answering this question

HOW BIG SHOULD MY SAMPLE BE?

Answer is …

$$n = \left[\frac{4\sigma^2 (z_{1-\alpha/2} + z_{1-\beta})^2}{D^2} \right] \left[1 + \rho(m-1) \right]$$

QUESTIONS?

ID4D IDENTIFICATION FOR DEVELOPMENT

i2i DIME TRANSFORM DEVELOPMENT

BILL & MELINDA GATES foundation

THE WORLD BANK

UKaid from the British people

A better approach…

What influences the sample size (η) I need?

1. Minimum detectable effect size ($\mathcal{D}$)
   - Statistical power (β) and confidence (α)
2. Variation in outcome (σ)
3. Clustering (η, ρ)
4. Take-Up
5. Data Quality

effect size

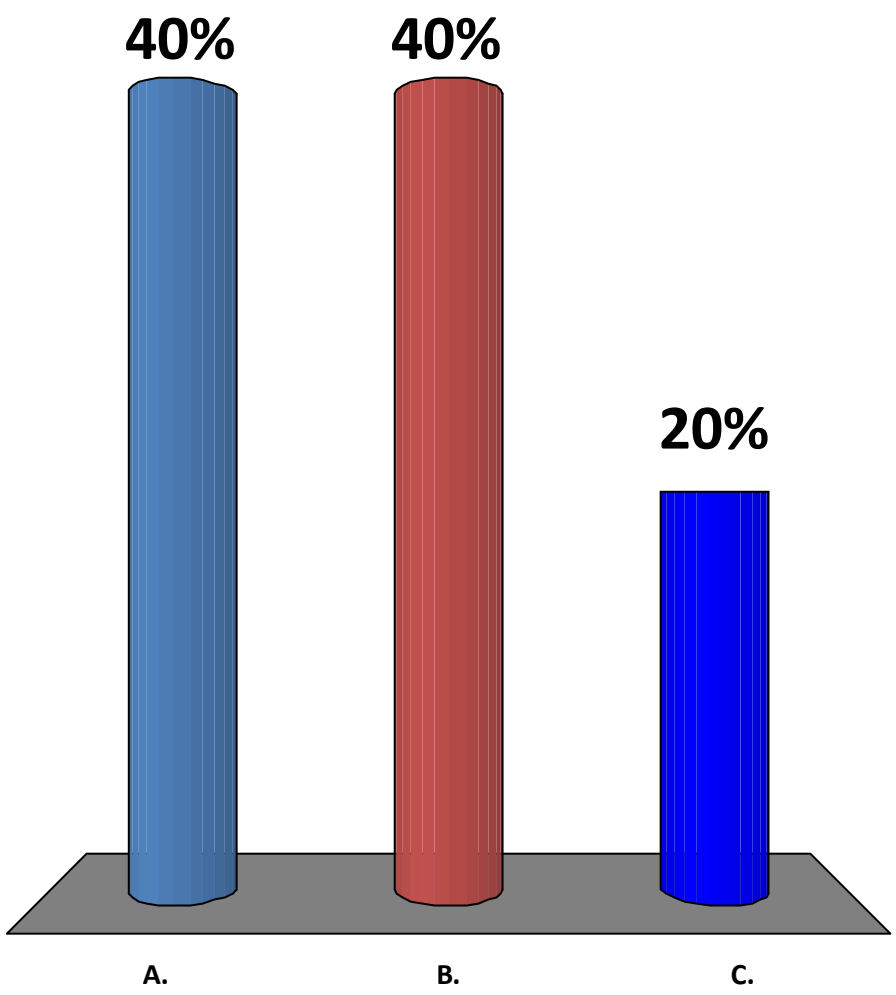
$$n = \left[\frac{4\sigma^2 (z_{1-\alpha/2} + z_{1-\beta})^2}{D^2} \right] \left[1 + \rho(m-1) \right]$$

Which IE will likely require a larger sample?

A. An IE of a project expected to increase household income by at least 50%

B. An IE of a project expected to increase household income by at least 5%

C. Sample should be the same for both



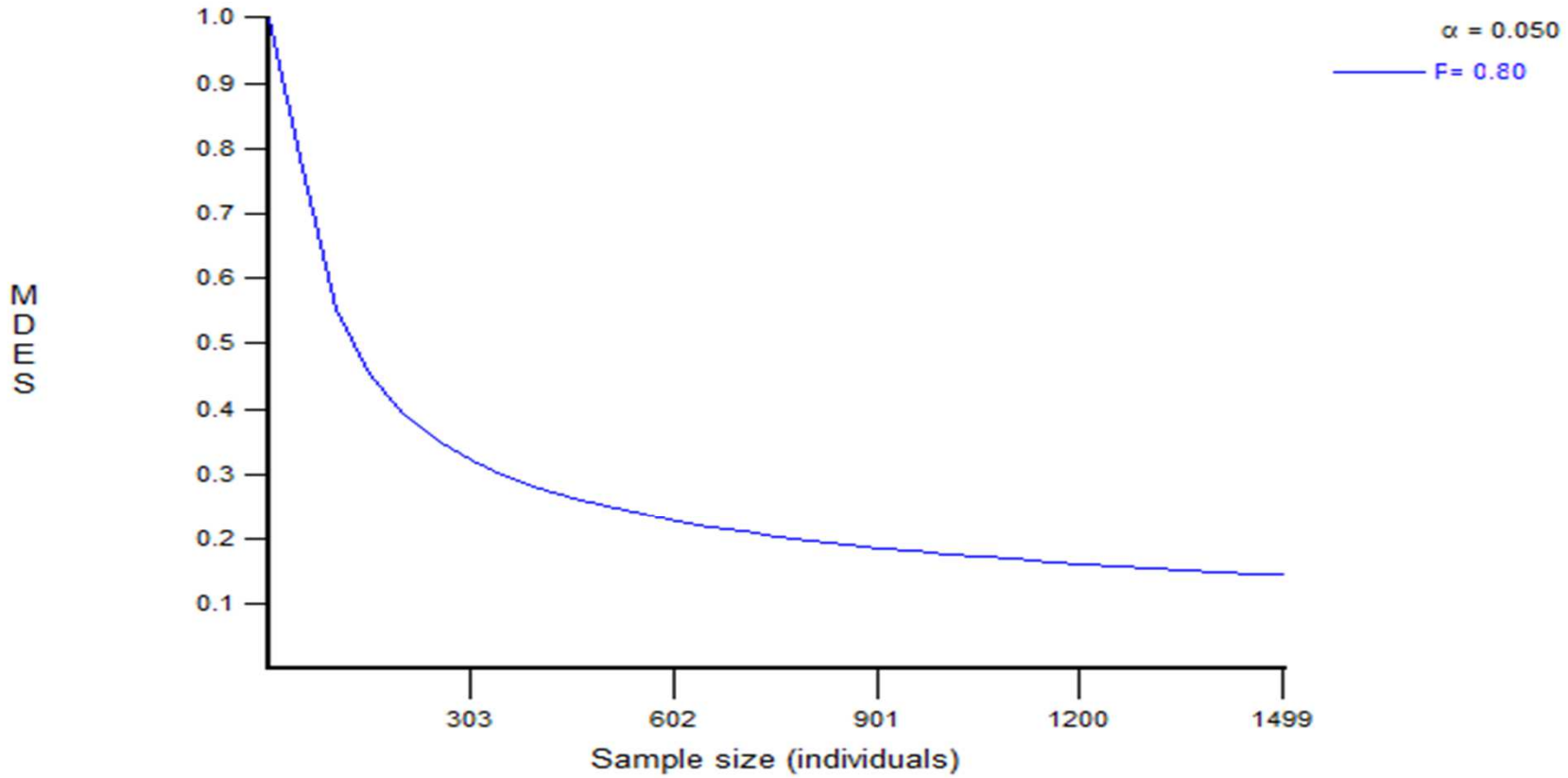
expected impact

- *"What is the smallest effect size that, if it were any smaller, the intervention would not be worth the effort?"*
  - Called **Minimum Detectable Effect Size** (MDES)
- The smaller the effect you want to be able to detect, the larger the sample you will need
  - larger sample ➔ more precise measuring device

expected impact

Increasing the sample acts as a magnifying glass to improve precision

expected impact



power and confidence

$$n = \left[\frac{4\sigma^2 (z_{1-\alpha/2} + z_{1-\beta})^2}{D^2} \right] \left[1 + \rho(m-1) \right]$$

power & confidence

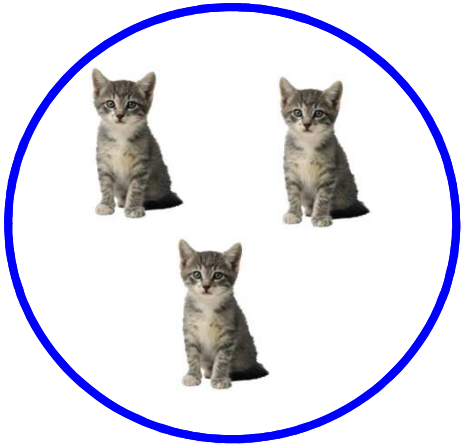
- Statistical confidence
  - likelihood of type 1 error → reject null when true
  - Standard assumption: $\alpha = .05$
- Statistical power
  - likelihood of type 2 error → fail to reject null when false
  - Standard assumption $\beta = 80\%$

variance of outcomes

$$n = \left[\frac{4\sigma^2 (z_{1-\alpha/2} + z_{1-\beta})^2}{D^2} \right] \left[1 + \rho(m-1) \right]$$

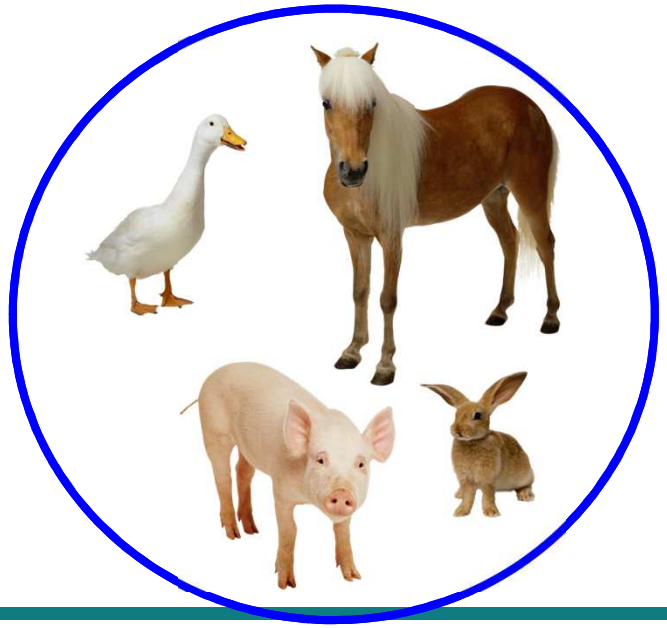
variance of outcomes

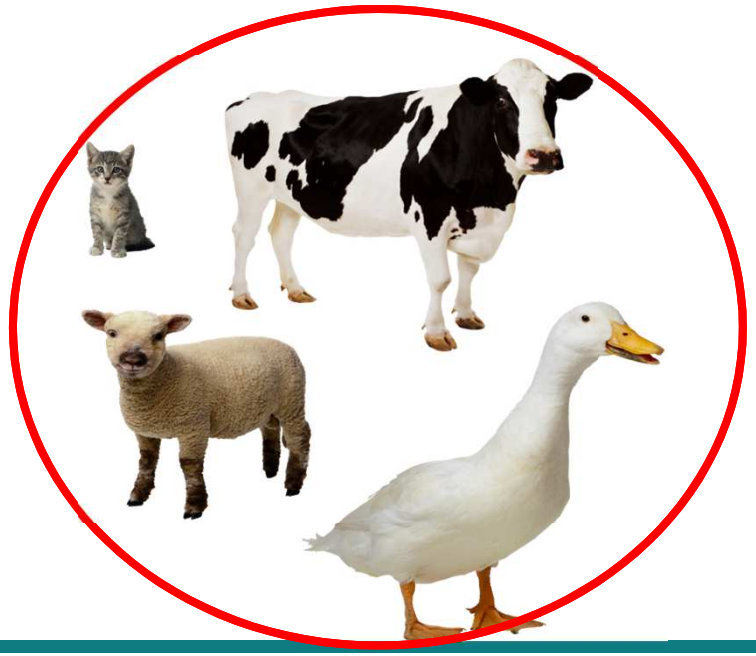
- Of the two (circled) populations, which animals are bigger?
- How many observations from each would you need to decide?



variance of outcomes

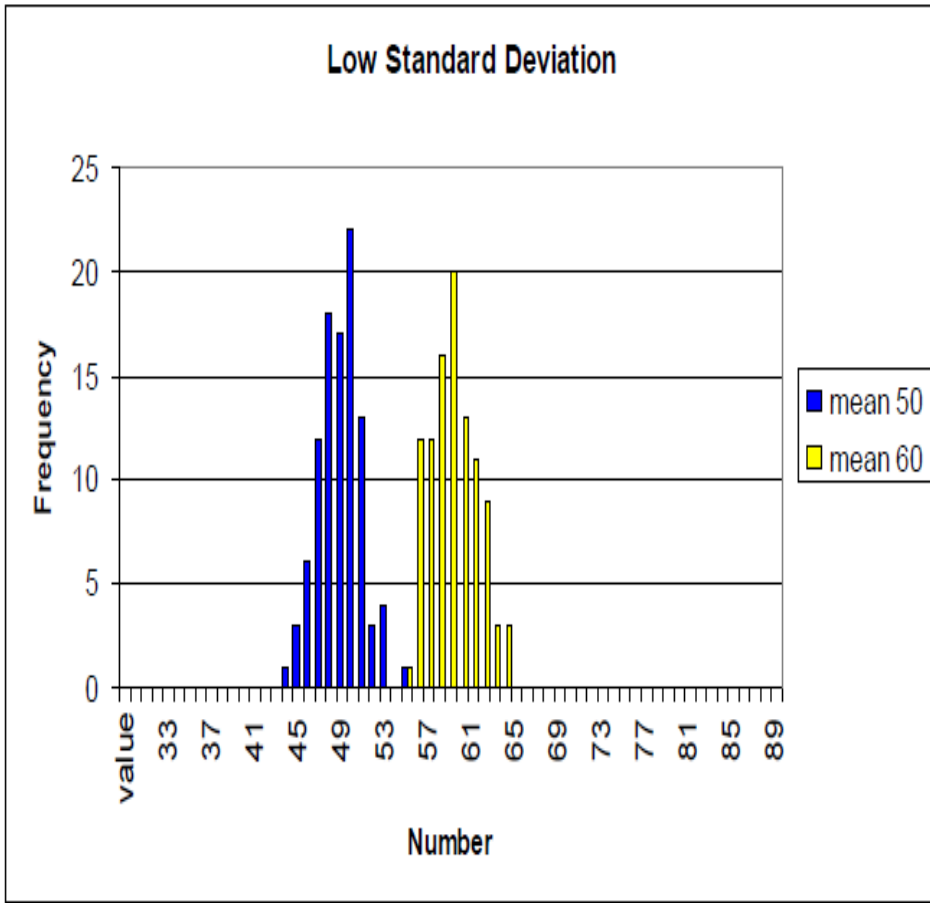
- Of the two (circled) populations, which animals are bigger?
- How many observations from each would you need to decide?

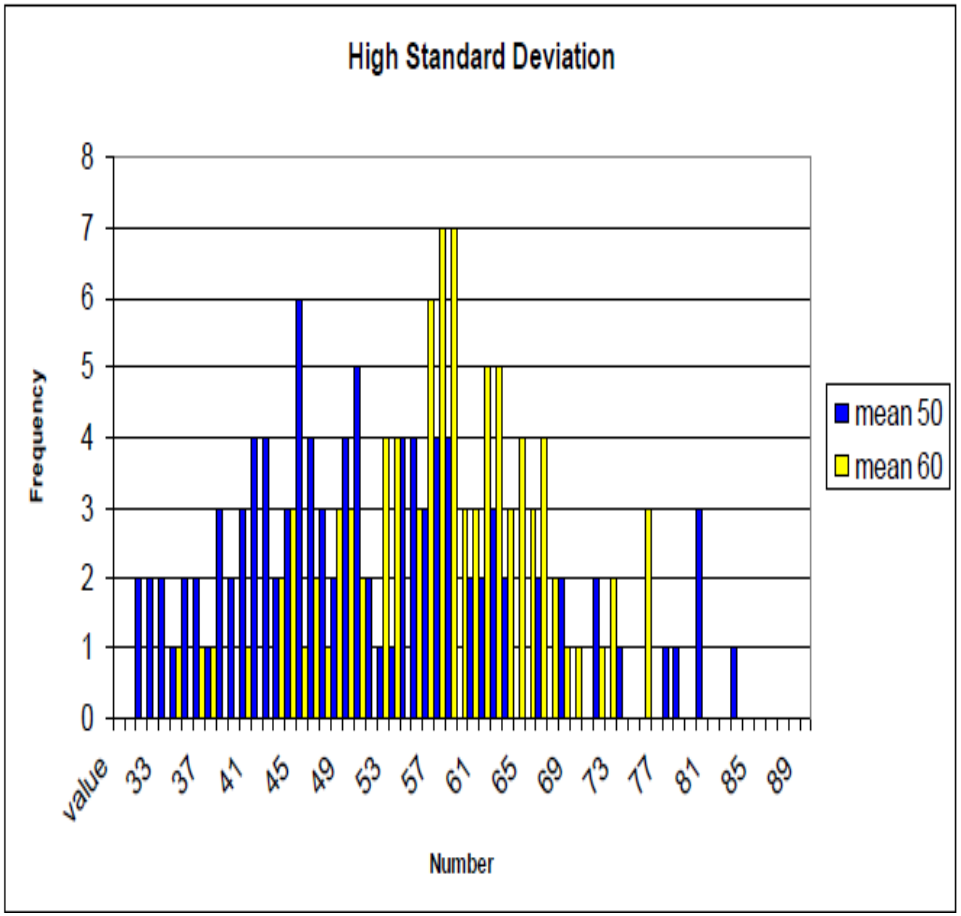




QUIZ

- A subsidy increases employment by 10% for the treatment group on average, in both cases below

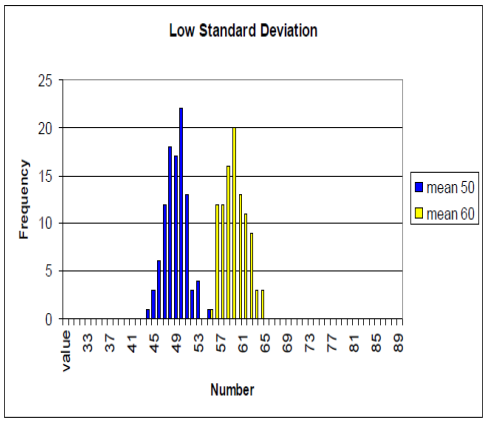


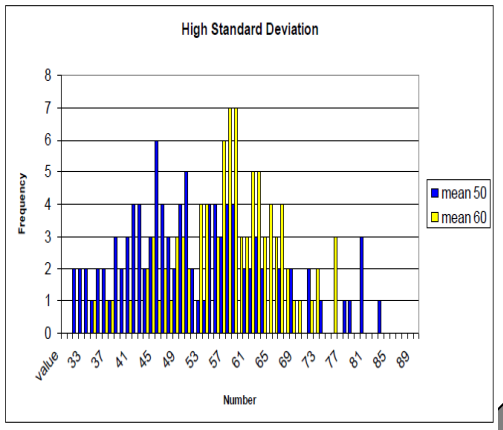


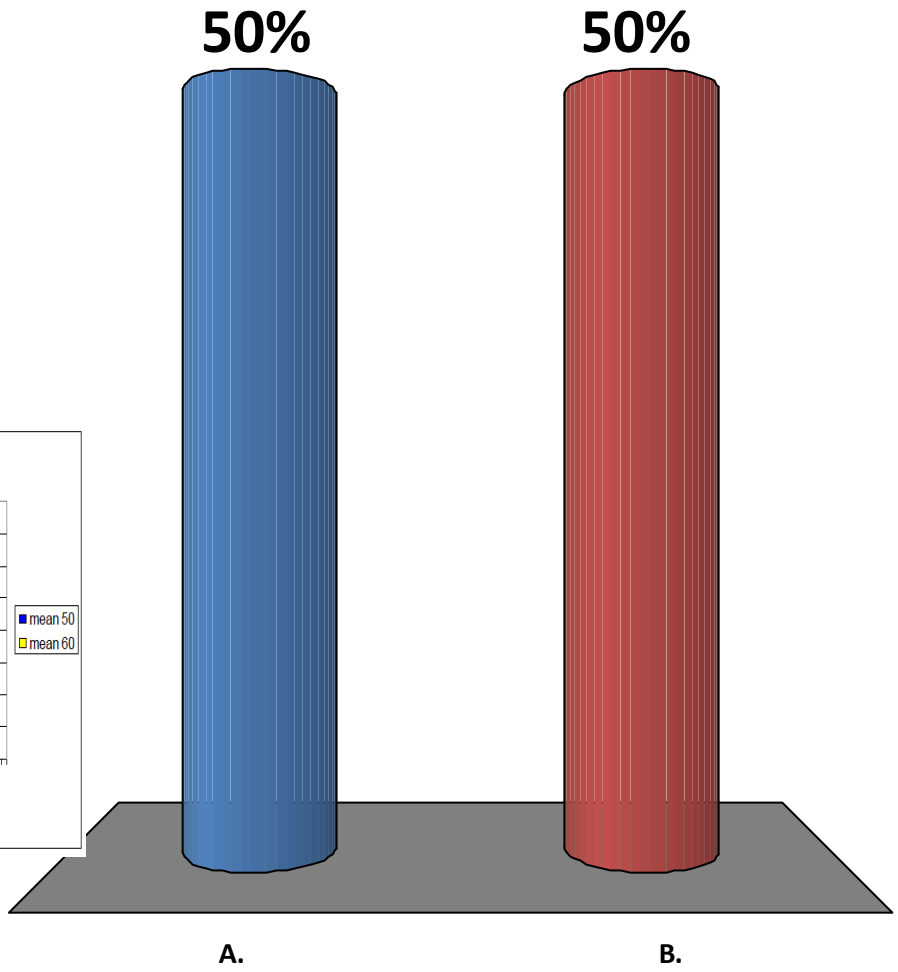
Which case requires a larger sample?

A. Low standard deviation case

B. High standard deviation case







variance of outcomes

- **In sum:**
  - More underlying variance (heterogeneity)
  - → more difficult to detect difference
  - → need larger sample size

- **Tricky:** How do we know about heterogeneity *before* we decide our sample size and collect our data?
  - Ideal: pre-existing data … but often non-existent
  - Can use pre-existing data from a *similar* population
    - Example: LSMS, data routinely collected by govt, satellite imagery
  - Common sense

clustering (aka “design effect”)

$$n = \left[\frac{4\sigma^2 (z_{1-\alpha/2} + z_{1-\beta})^2}{D^2} \right] \left[1 + \rho(m-1) \right]$$

clustering

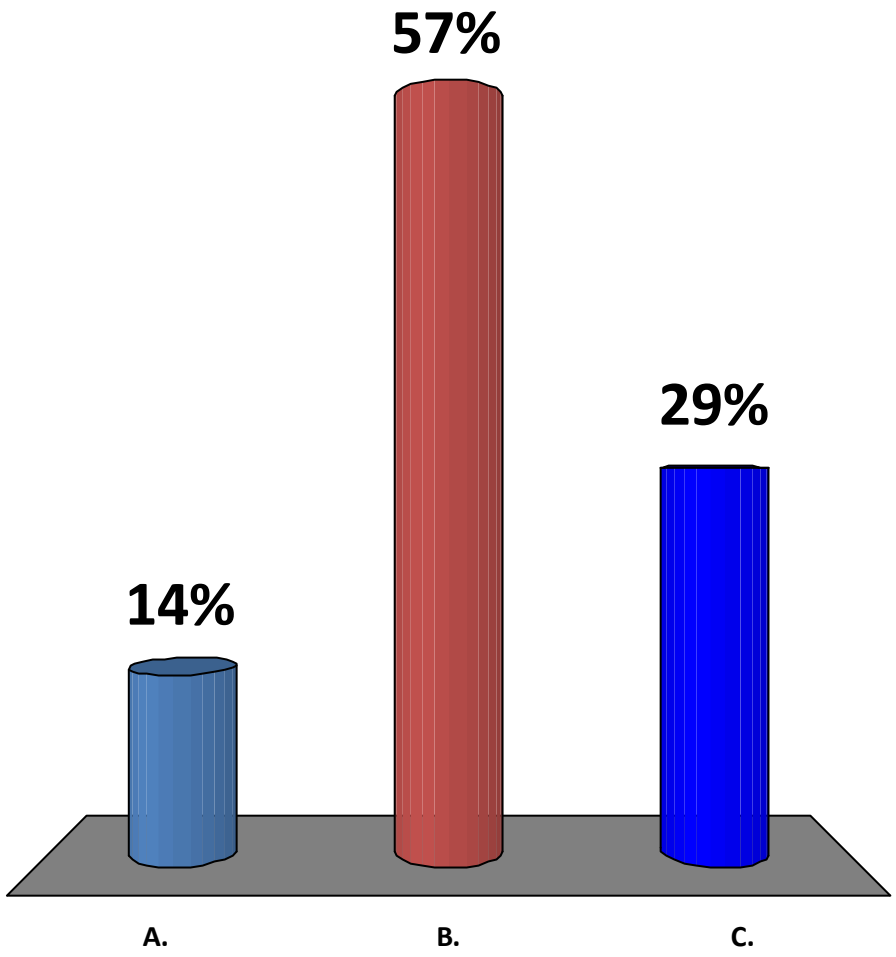
- Unit for sample size calculation depends on both:
  - Level of intervention AND
  - Level of measured impacts
- Example: intervention at village level, interested in impacts at HH level
  - Randomly assign villages to treatment / control
  - Sample household within villages

Which sampling strategy is likely to give you more statistical power?

A. 400 villages, 5 HHs per village = 2,000 HHs

B. 50 villages, 40 HHs per village = 2,000 HHs

C. Both should give you similar statistical power



clustering

- Level of intervention ("cluster") most important for sample size calculation

- If few clusters, precision will be limited, regardless of number of HHs sampled

clustering

- Ex. Randomize transport voucher at village level, in 6 villages. Sample 1,000 HHs per vlg.

- Sample size: 6,000 HHs – that's a lot, right?!!
  - Key sample size number is 6
  - Adding clusters is **always** a better way to increase precision than adding HHs within clusters
  - How much precision the 1,000 HHs buys you depends on "intra-cluster correlation"

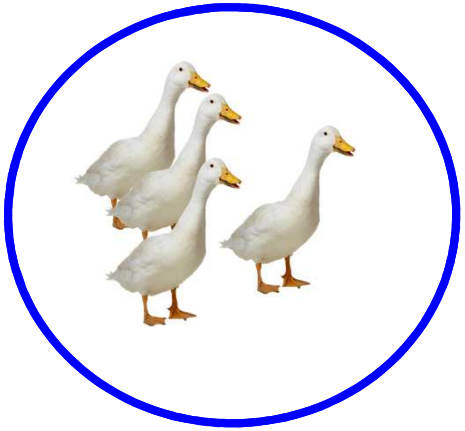
clustering

- **Intracluster correlation (ICC):** similarity of units within clusters

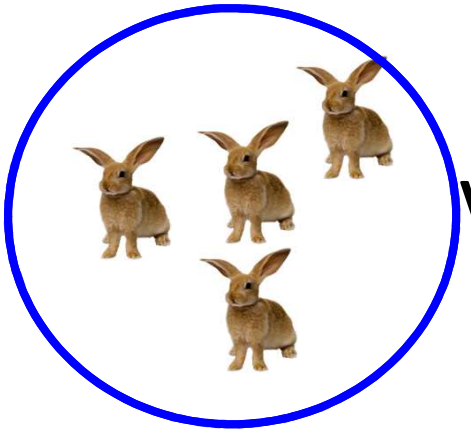
- Is the variation in outcome of interest coming mostly from differences *within* villages (low ICC), or *between* villages (high ICC)?
  - If HHs in village A are similar to each other, but different from HHs in village B, high ICC
  - If HHs in village A are similar to HHs in village B, low ICC

- If ICC = 0, no design effect

Clustering (high ICC)

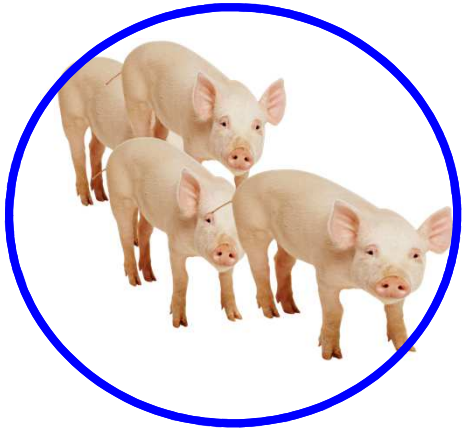


Village 1



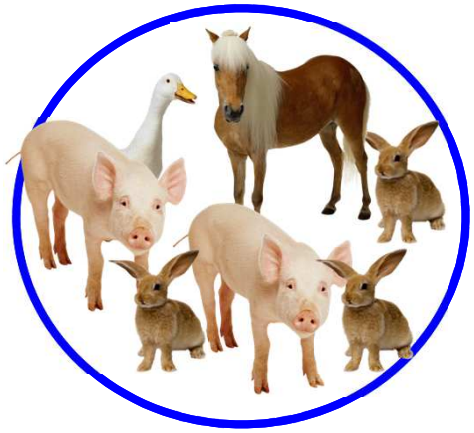
Village 3

Village 2



Village 4

Clustering (low ICC)



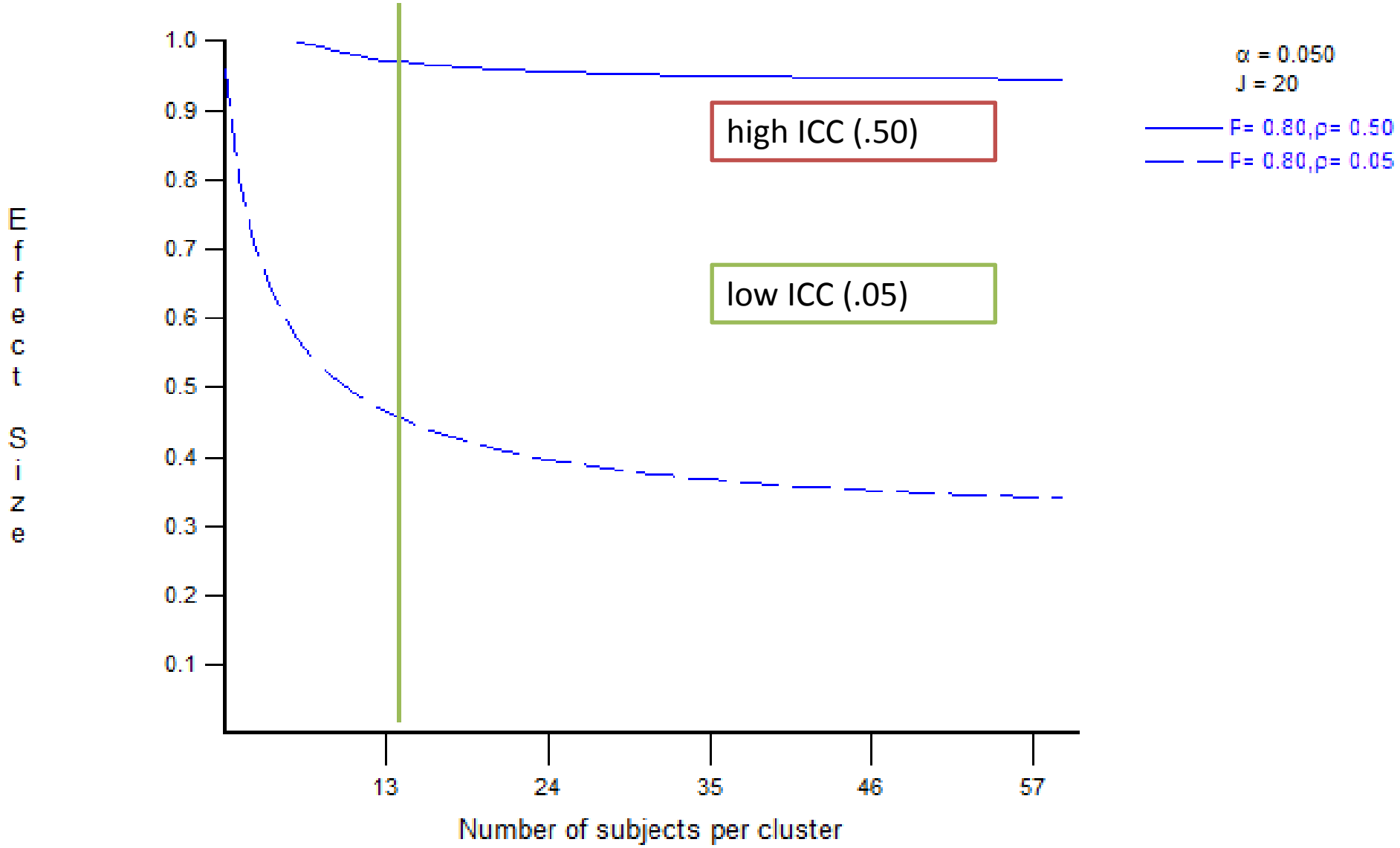
Village 1

Village 2

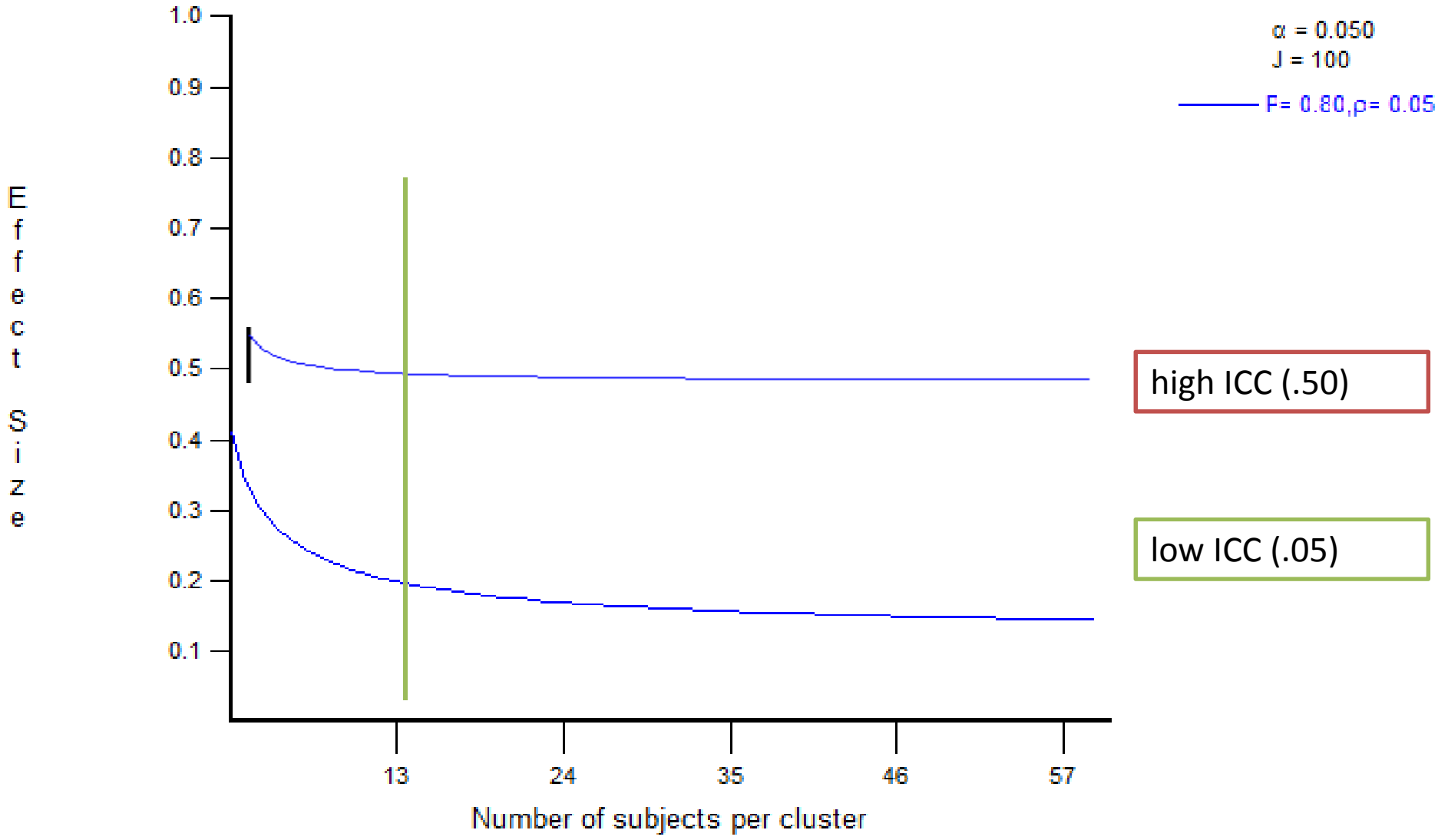
Village 3

Village 4

20 clusters


α = 0.050
J = 20
P= 0.80,ρ= 0.50
P= 0.80,ρ= 0.05
high ICC (.50)
low ICC (.05)
Effect Size
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
13
24
35
46
57
Number of subjects per cluster

100 clusters



clustering

Takeaway

High *intra-cluster correlation* (HHs in same cluster similar)

lower marginal value per extra sampled unit in the cluster

More clusters needed

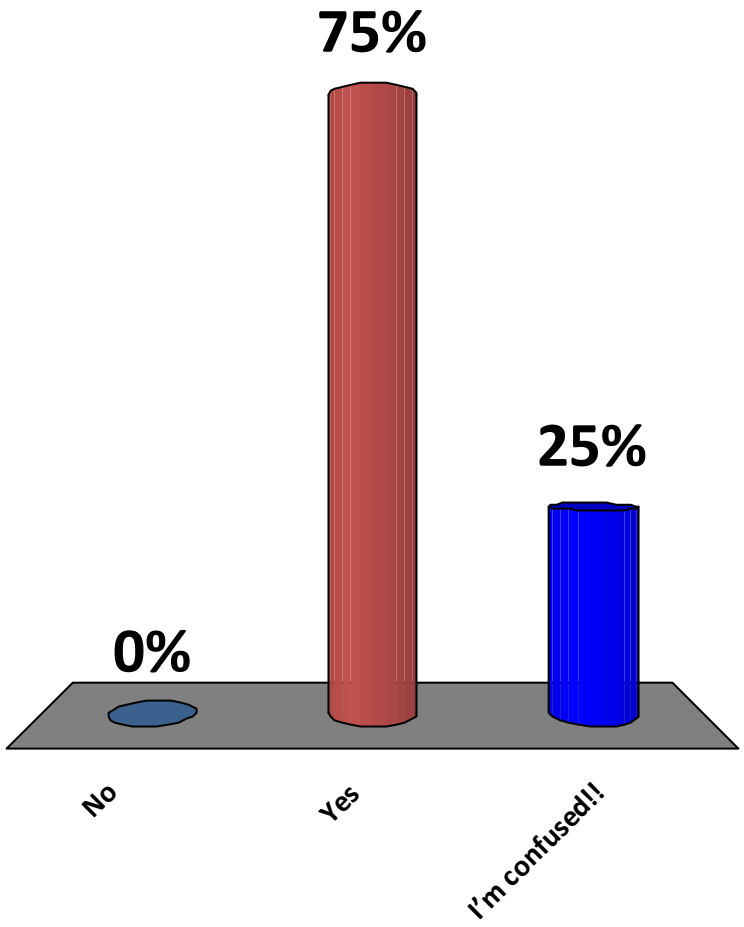
Rule of thumb: <u>at least</u> 40 clusters per treatment arm

take-up

- **Example: IE of a smart ID program**
  - You design a study to measure impact of ‘smart’ IDs.
  - Sample size calculations show you will need 1000 HHs in your study (500 treatment, 500 control).
  - You do a baseline survey of the 1000 HHs, then offer the smart ID to the 500 treatment households.
  - 250 of the treatment HHs decide to adopt the smart ID

Do you need to worry about statistical power for this IE?

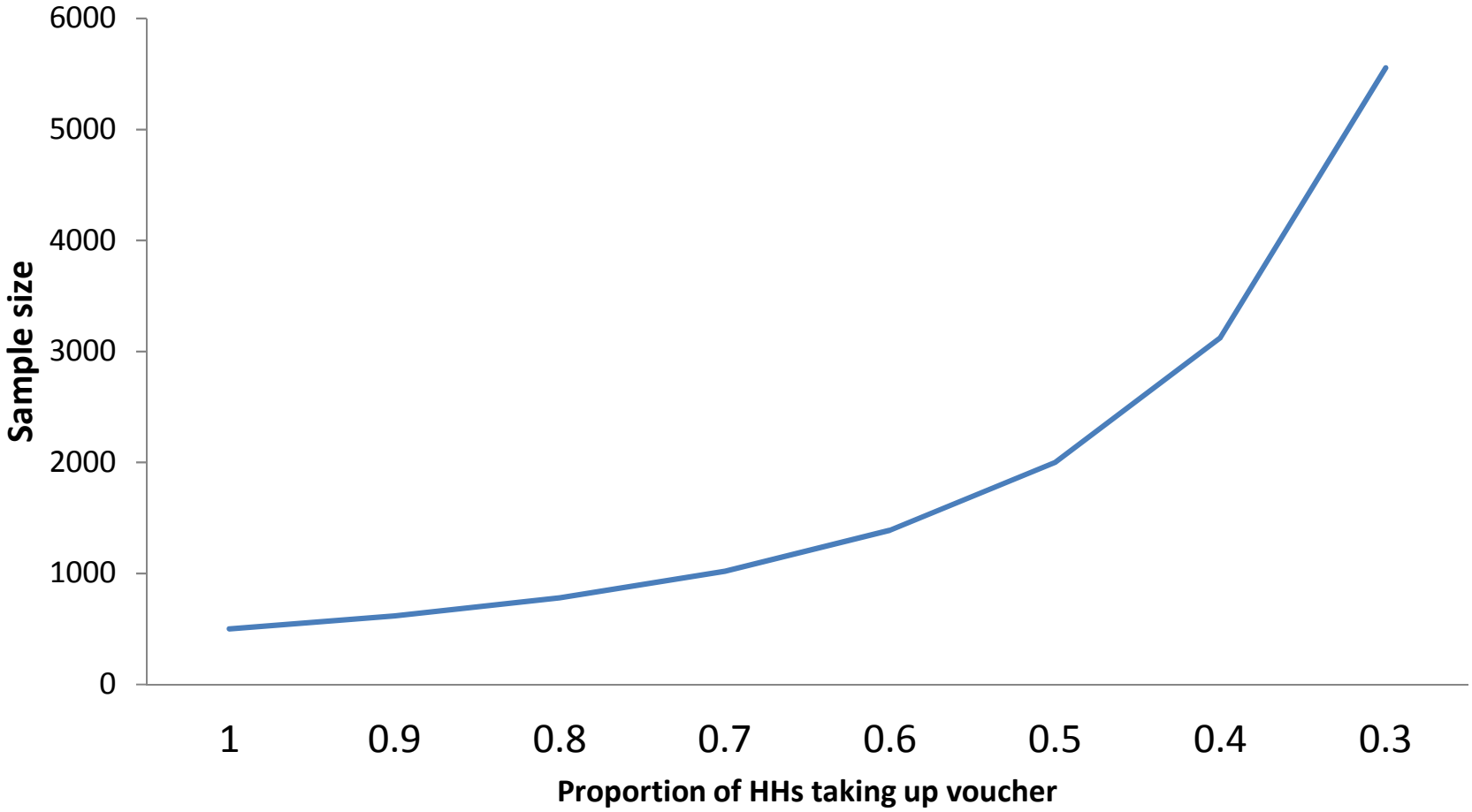
A. No
B. Yes
C. I'm confused!!



take-up

- Low take-up (rate) for intervention lowers precision
- Effectively decreases sample size / increases minimum detectable effect
  - Can only detect an effect if it is really large
- Unfortunately, to account for take-up rate of 50%, have to increase sample size by factor of 4

Take up vs. sample size



data quality

- Poor data quality effectively increases required sample size
  - Missing observations
    - ✓quality of data collection, attrition, migration
  - High measurement error: answers not always precise
    - ✓e.g. self-reported land size, agricultural production
    - ✓e.g. recall bias, framing, pleasing

- Poor data quality can be **partly** addressed with field coordinator on the ground **monitoring data collection**

conclusions

The smaller effects that we want to detect	**The larger the sample size has to be**
The more underlying heterogeneity (variance)	
The higher the level of clustering	
The lower take up	
The lower data quality	

To keep in mind this week

- What is the …
  - level of randomization (clustering)?
  - Expected effect size?
  - Variation within target population?

- How to ensure …
  - High take-up?
  - Good data quality?

If you like the graphs you saw here…

- You can make your own with Optimal Design, a free download from Univ. of Michigan

http://sitemaker.umich.edu/group-based/optimal_design_software