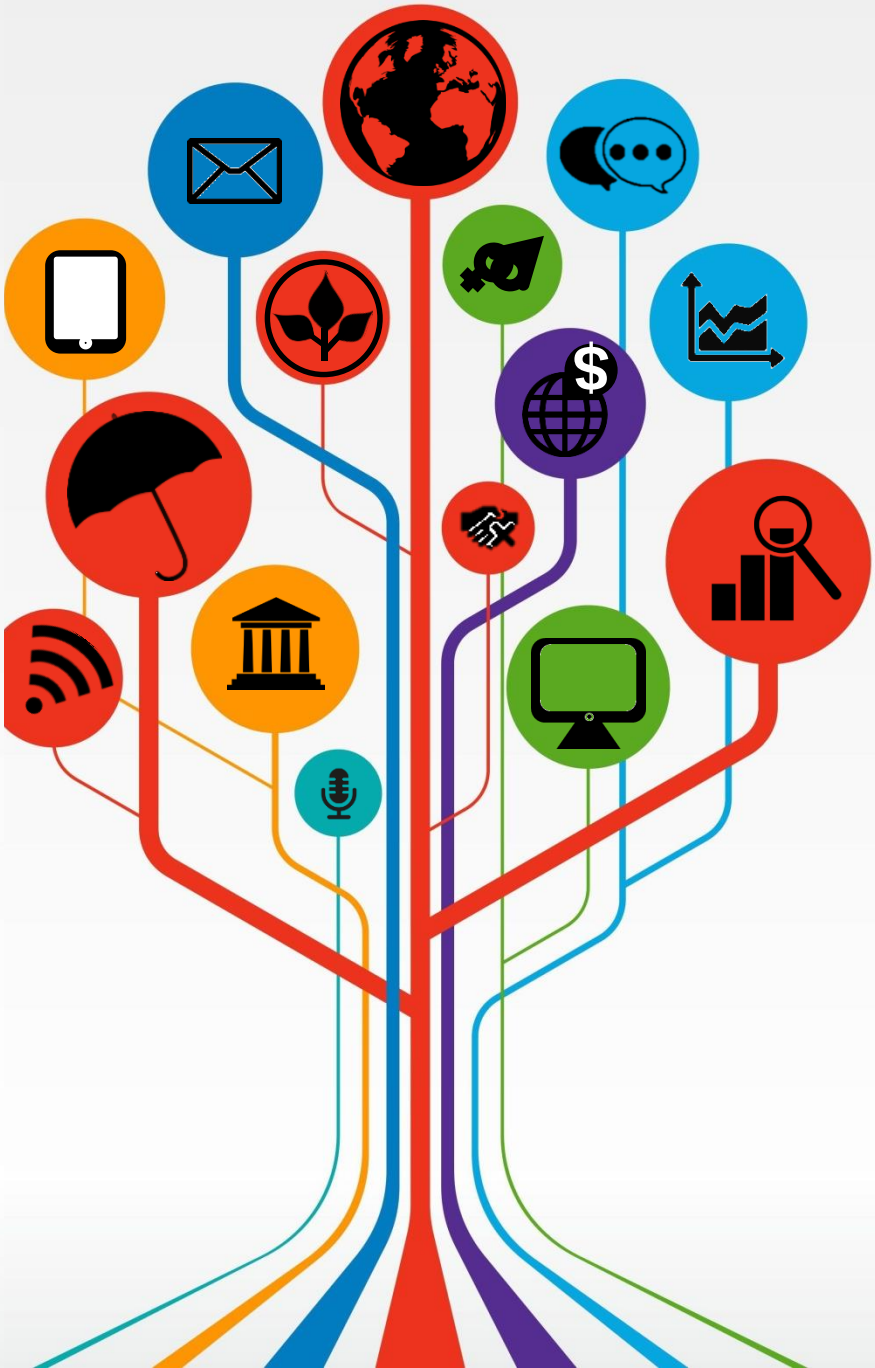
FIELD COORDINATOR WORKSHOP

Manage Successful Impact Evaluations

10-14 JUNE 2019
WASHINGTON, DC



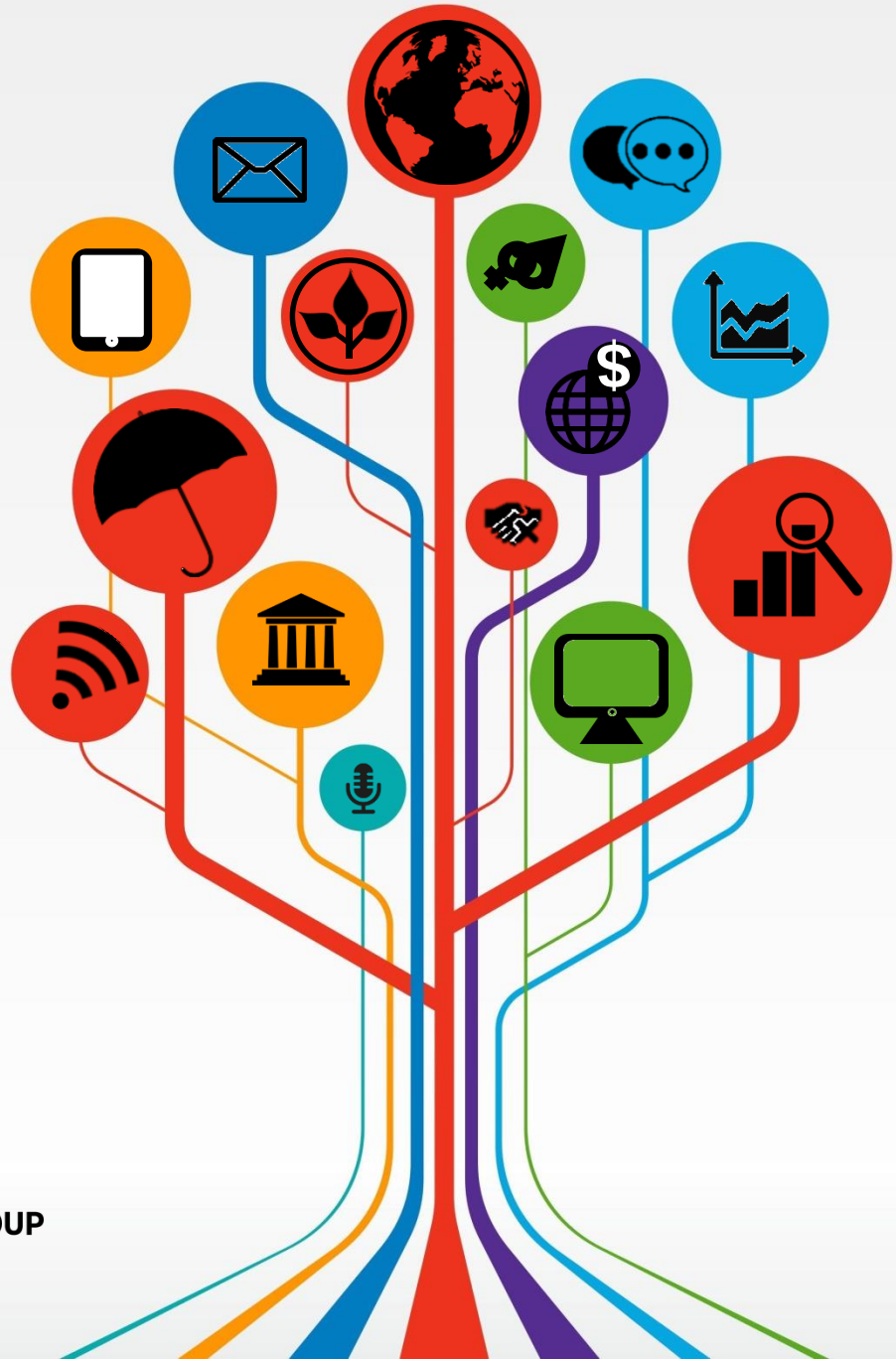
i2i DIME
TRANSFORM DEVELOPMENT

WORLD BANK GROUP

Sampling for people that love sampling things.

Aidan Coville

13 June, 2019



i2i DIME
TRANSFORM DEVELOPMENT

WORLD BANK GROUP

SAMPLING MYTHBUSTERS

Is a sample always representative of the population?

No WAYS!

Is random sampling the same as random assignment of an intervention?

No WAYS!

Can sampling be fun?

No WAYS!

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

This presentation covers two questions:

Why is sample size important?

- Approx time: 2 mins

How big should my sample be?

- Approx time: a lifetime of pain and anguish

Q1: Why is sample size important?

Why is sample size important?

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

	S					M		
T	H					N		

Why is sample size important?

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

	S	H		W		M		
T	H			M	O	N		Y

Why is sample size important?

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

	S	H	O	W		M	E	
T	H	E		M	O	N	E	Y

Why is it important for IE?

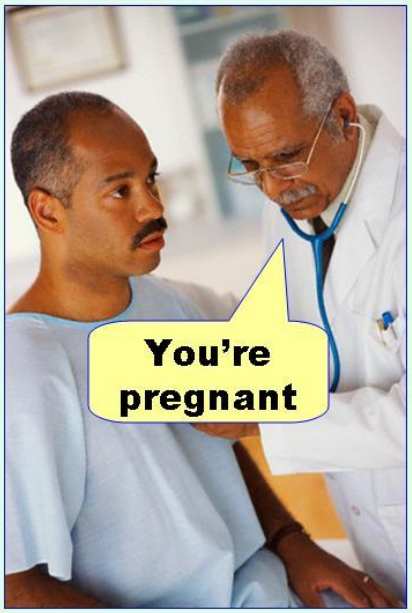
We want to know the true impact

But we need to estimate this impact from a sample

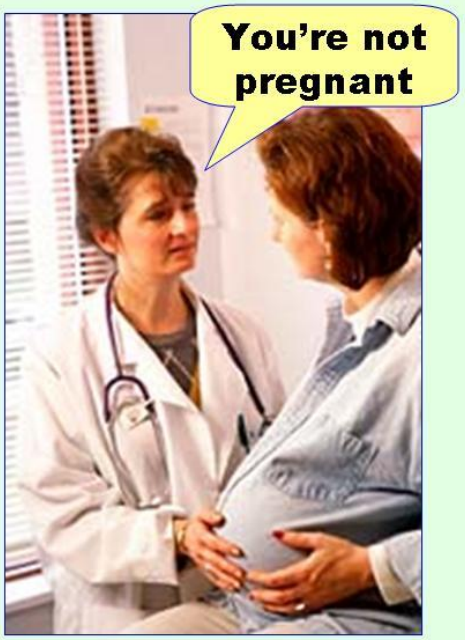
Estimation means we can sometimes make mistakes

Making mistakes can be costly…

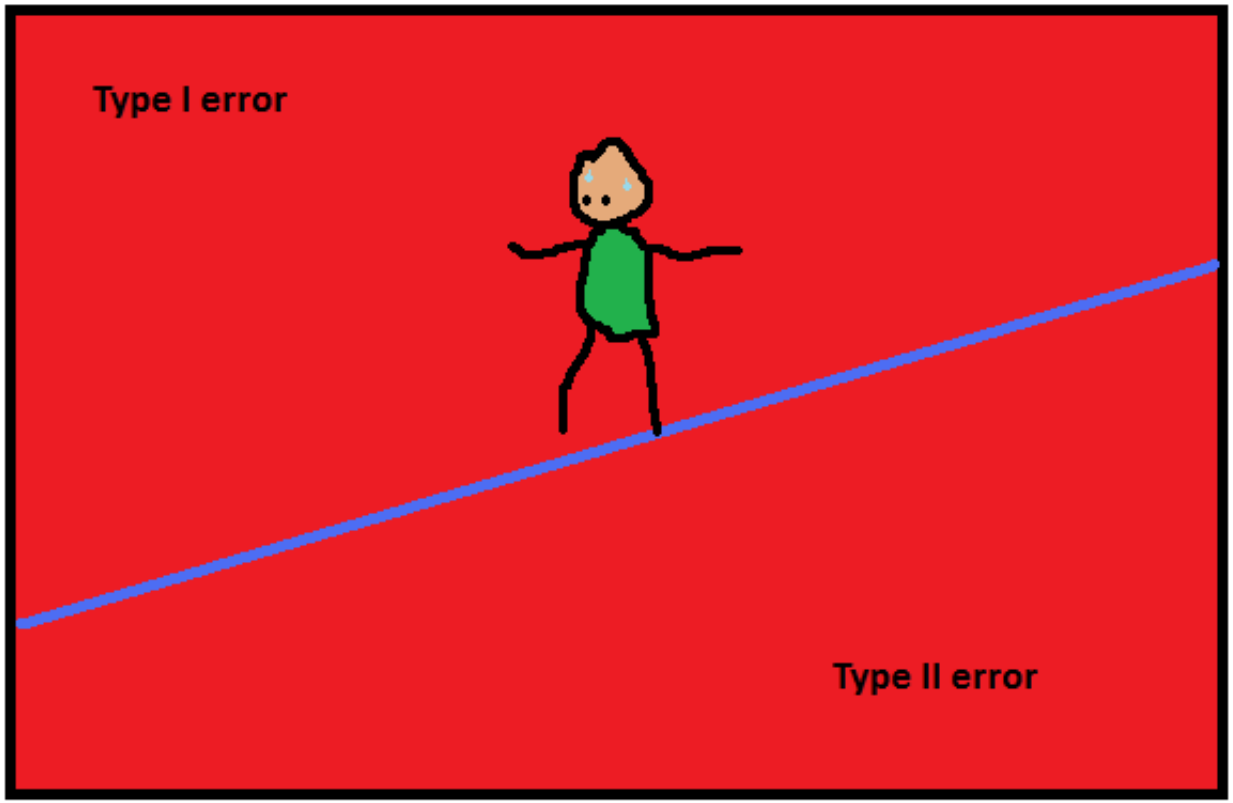
Type I error
(false positive)



Type II error
(false negative)



Balance between statistical significance and power



Q2: How big should my sample be?

The answer is…

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

= 42

The End

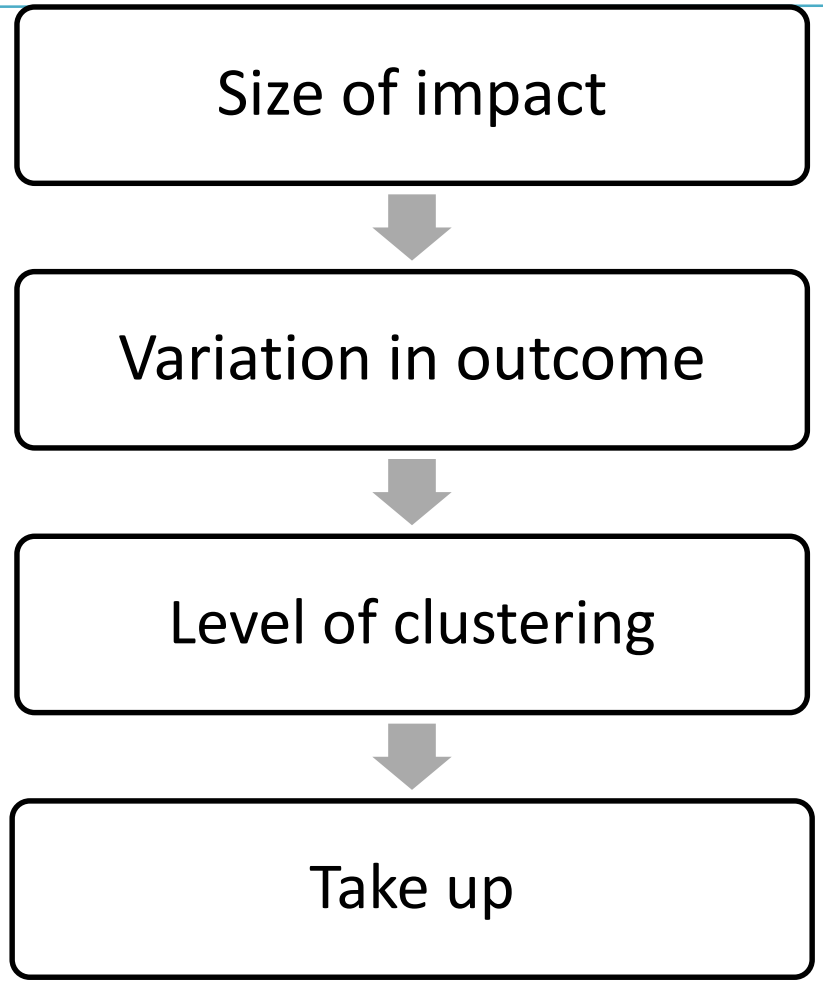
- Questions?

A better question…

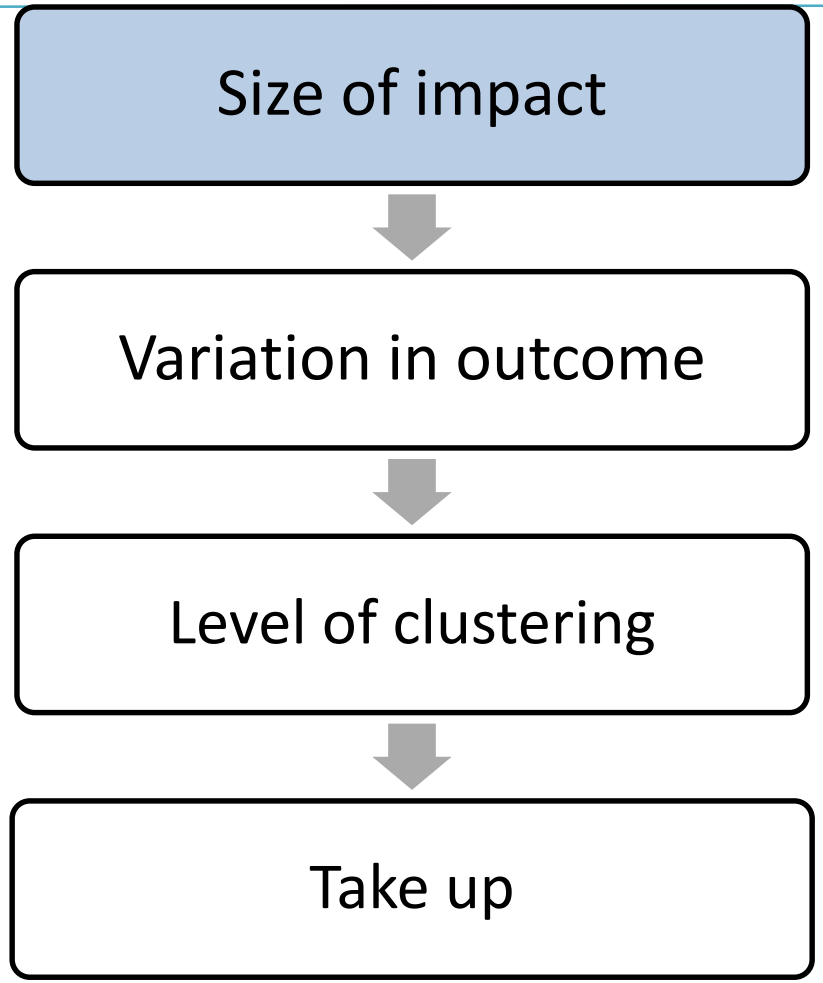
What influences the sample size I need?

- Size of impact
- Variation in outcome
- Level of clustering
- Take up

What influences the sample size I need?



What influences the sample size I need?



Size of impact

Big impacts are easy to identify

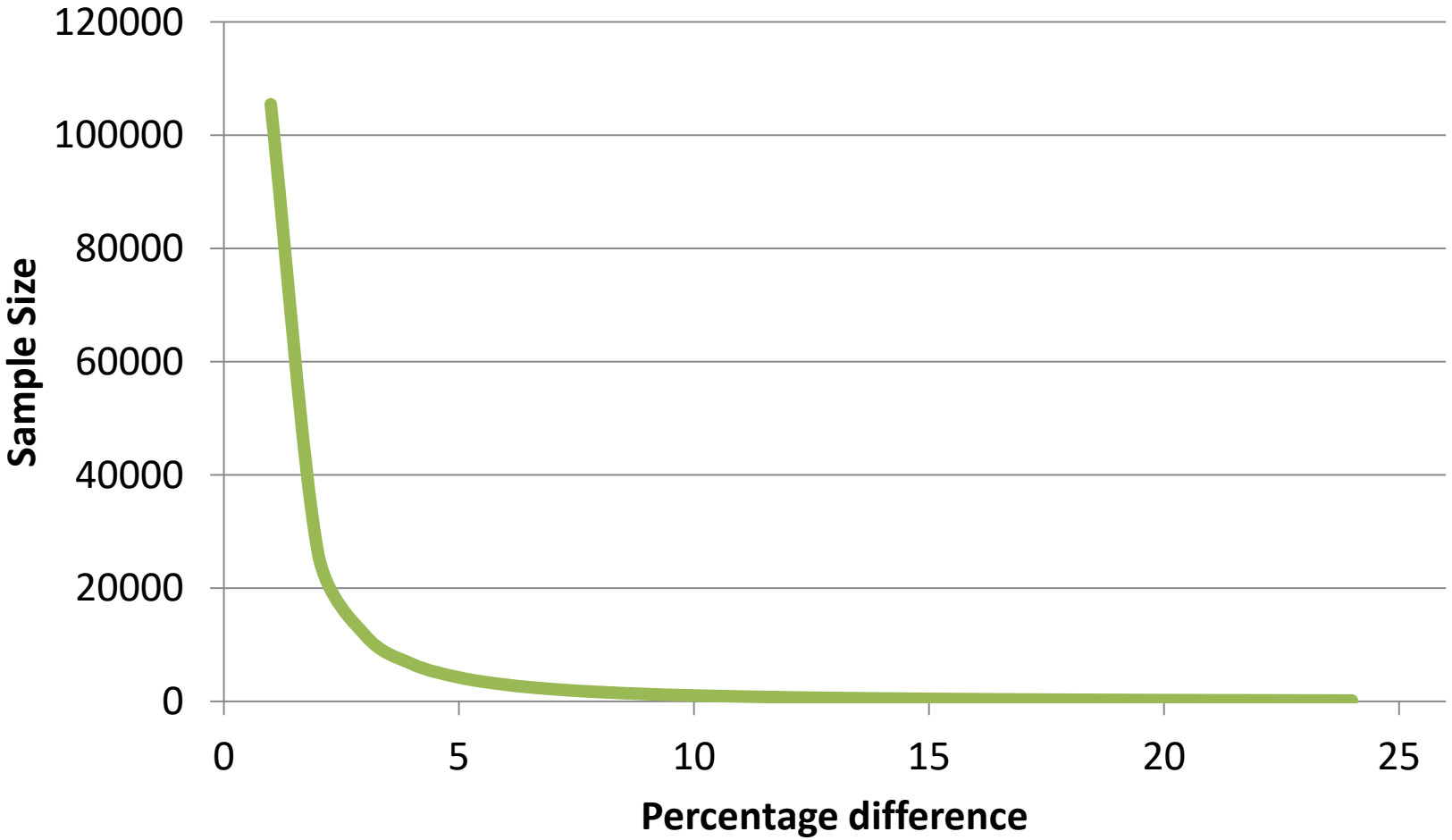
Small impacts are more difficult
Need more precision/accuracy
Larger sample needed

Minimum detectable effect

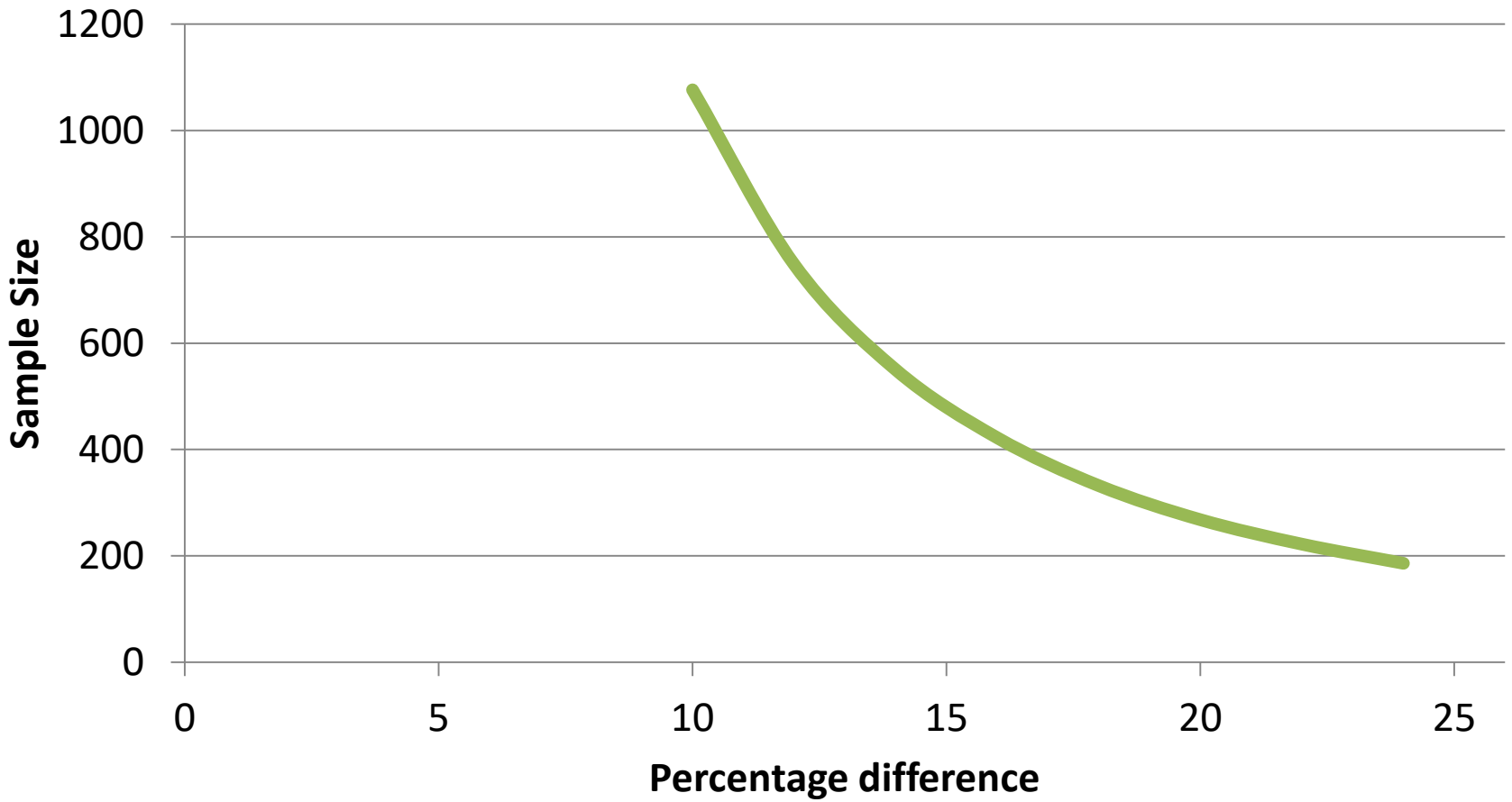
- We need a sample size able to detect the smallest effect size of importance.
- To guide this decision we need to ask:

"What is the smallest effect size that, if it were any smaller, the intervention would not be worth the effort?"

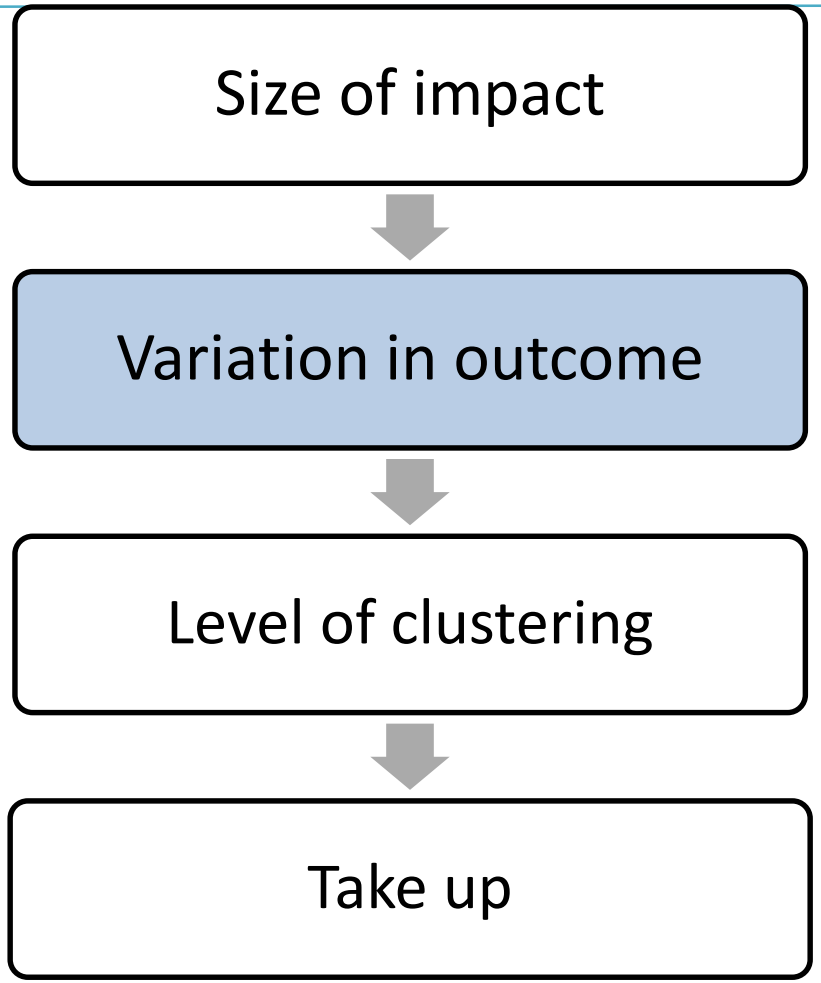
Mo money mo power



Need to be realistic

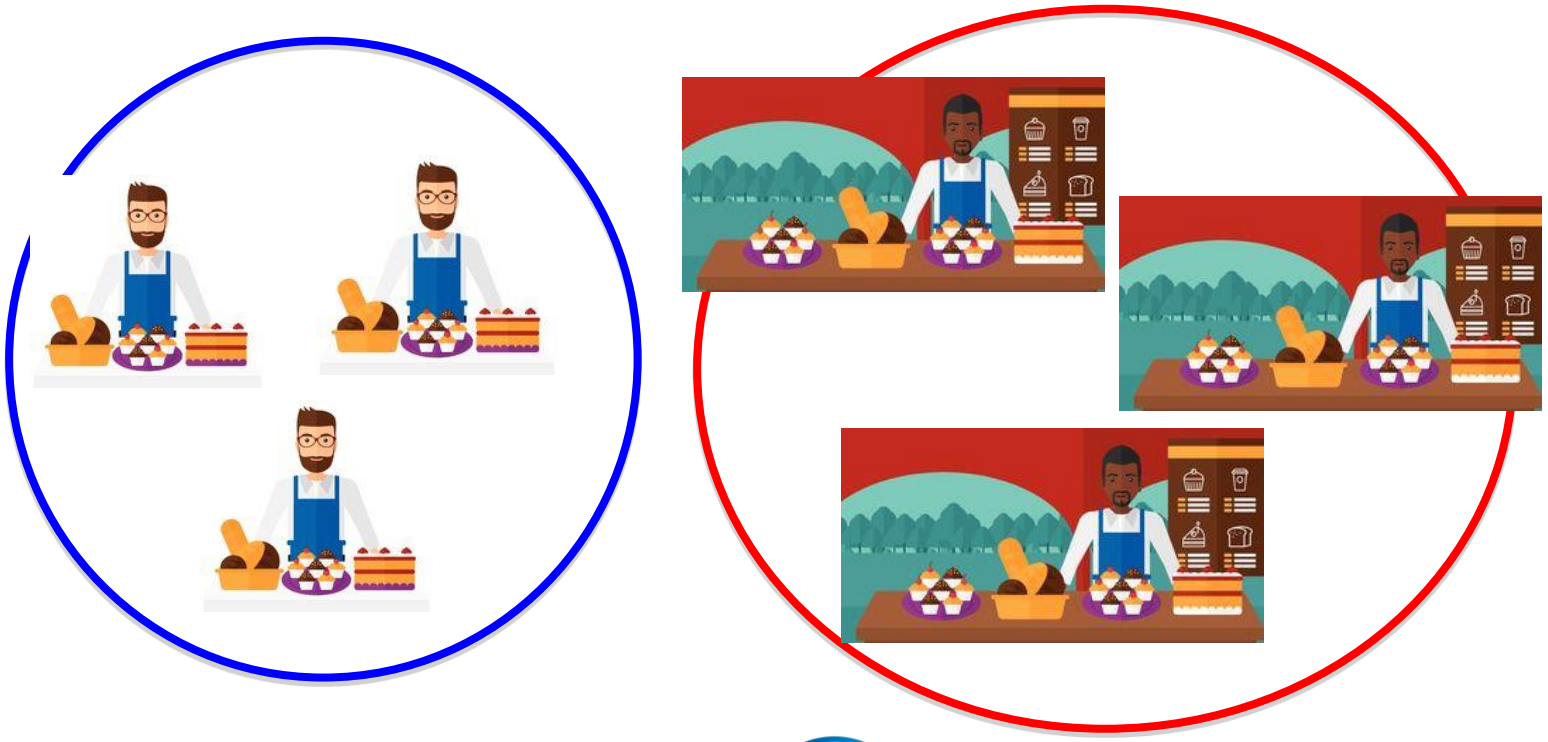


What influences the sample size I need?



Which group has more to sell?

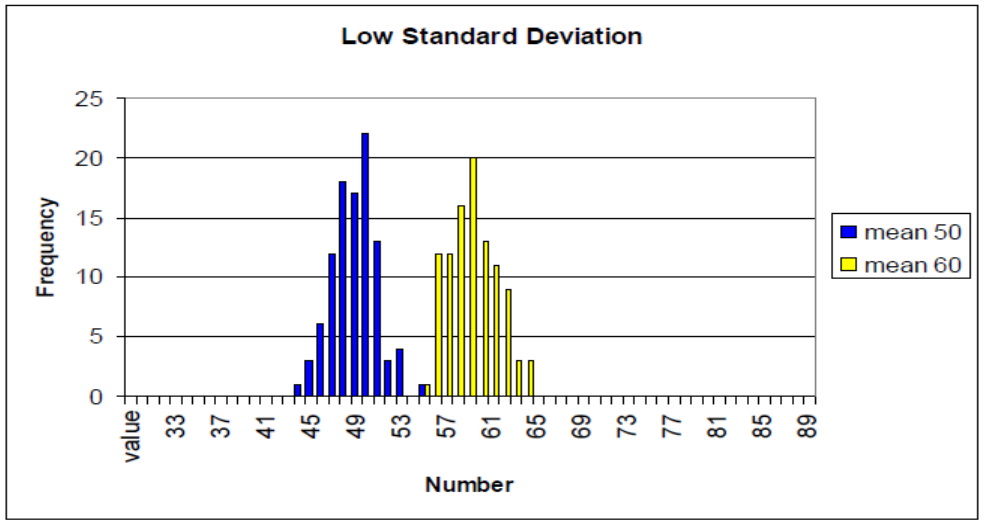
- How does the variance of the outcome affect our ability to detect an impact?

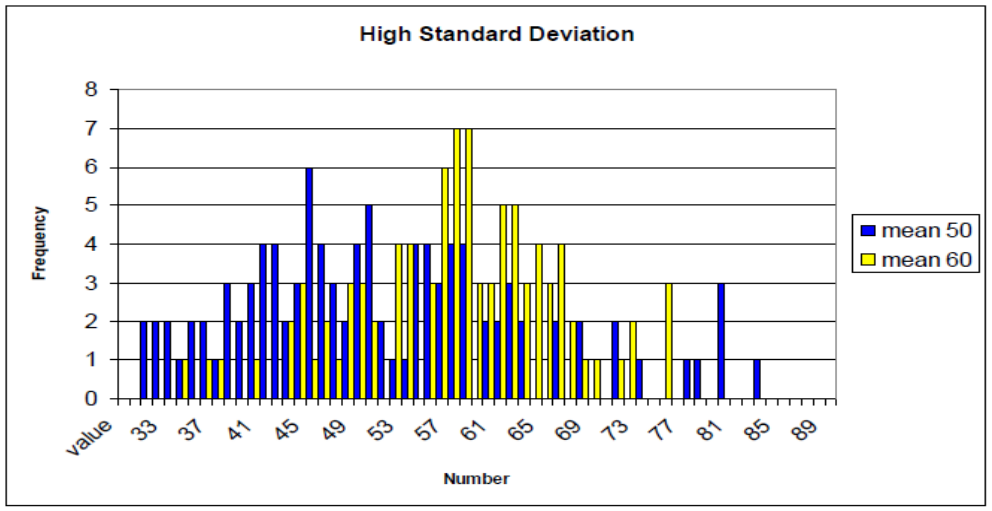


Now… which group has more to sell?

- How does the variance of the outcome affect our ability to detect an impact?

Which instance requires a larger sample?





Variation in outcomes (summary)

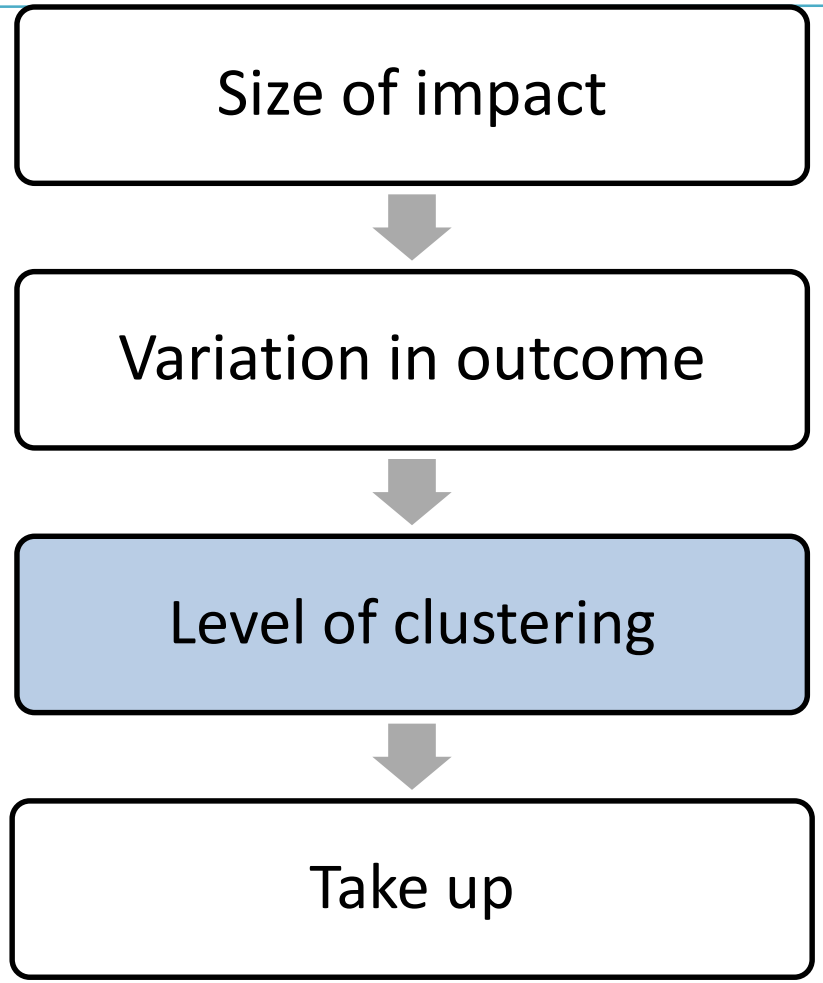
In sum:

- 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: enterprise surveys, labor force surveys
- Common sense

What influences the sample size I need?



Clustering (1/4)

For logistical or spillover reasons we may want to randomize at the group level.

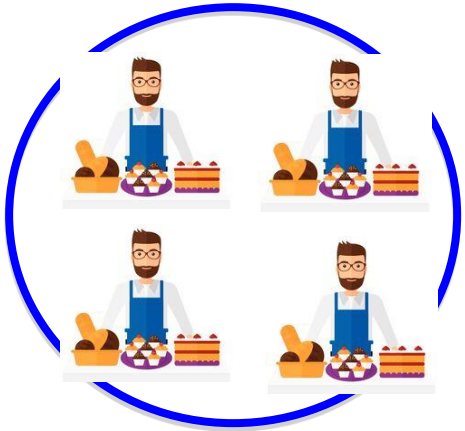
Sample size required increases, the higher the level of intervention assignment

- Business level
- Business group level
- Village/port/…
- Province?

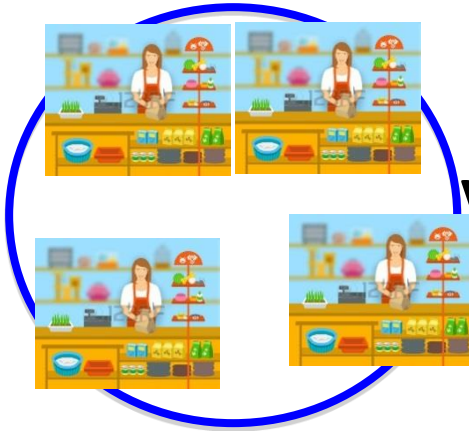
Even if unit of analysis is the firm/household/child, if level of randomization is at province (cluster) level, we run into challenges quickly…

Clustering (2/4)

What is the added value of more samples in the same cluster?

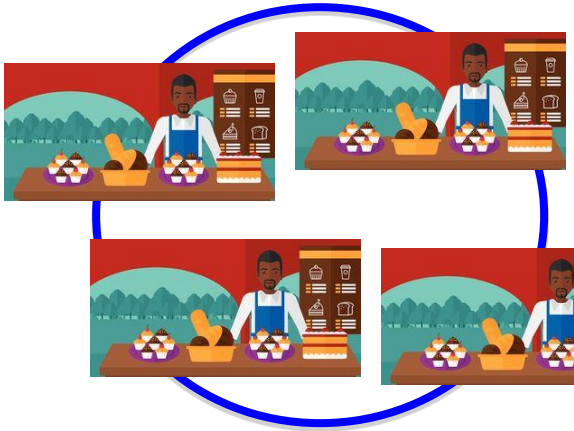


Village 1



Village 3

Village 2



Village 4

Clustering (3/4)

Village 1

Village 3

Village 2

Village 4

Clustering (4/4)

Takeaway

Larger *within cluster* correlation (guys in same cluster are similar)

lower marginal value per extra sampled unit in the cluster

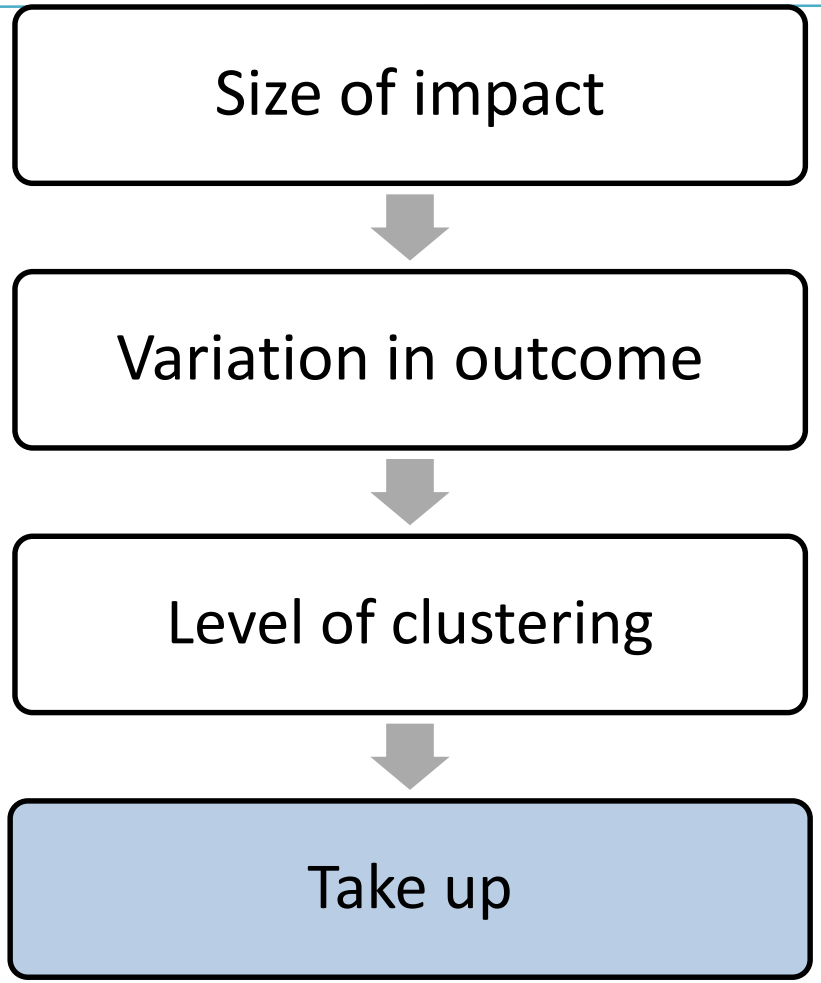
higher sample size/more clusters needed than a simple random sample.

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

For an intervention randomized at the village level, which sample design will have the most precision (highest power)?

A. 100 households from each of 2 villages
B. 25 households in each of 8 villages
C. 5 households in each of 40 villages

What influences the sample size I need?

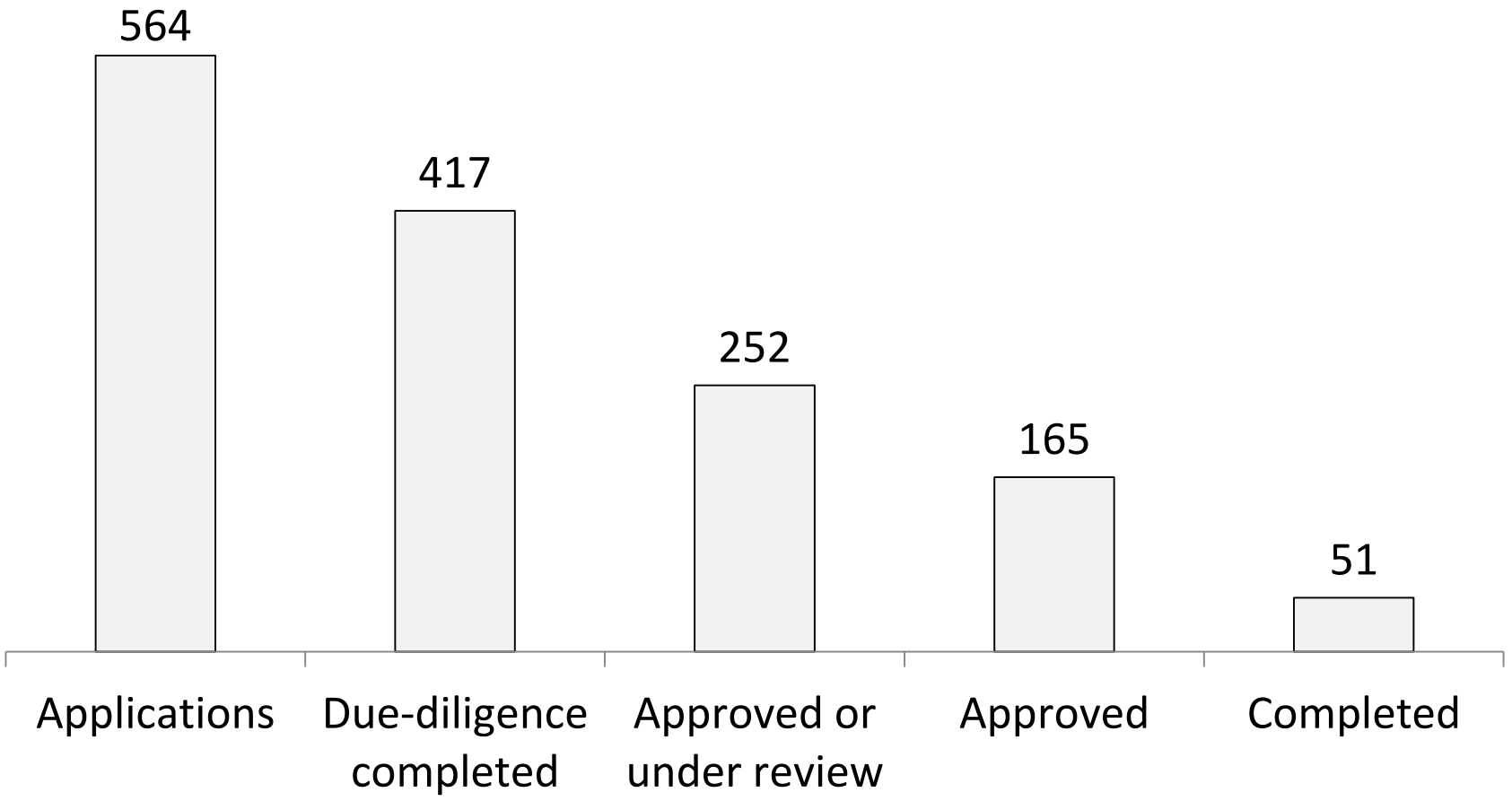



Oversubscription

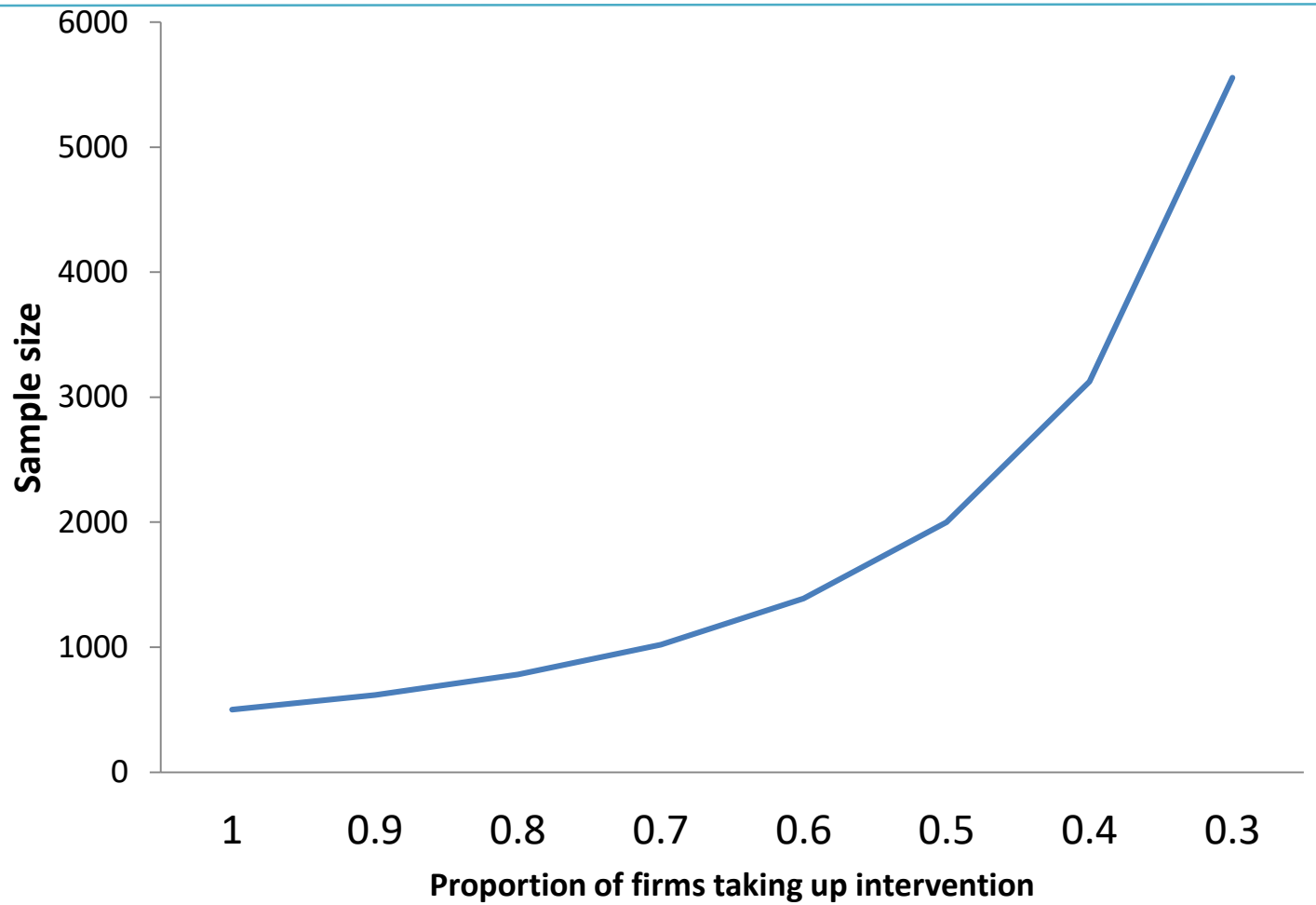

136
Over-optimistic

A real-life example

Matching grant application vs. completion rates



Take up vs. sample size



If my study has 50% take up, I will need a sample X times larger than if we had full compliance

A. 2x
B. 4x
C. 8x
D. 20x
E. 100x

Overview

- **Who** to interview is ultimately determined by our research/policy questions
- **How Many**:

Elements:	Implication for Sample Size:
The **smaller effects** that we want to detect	The **larger** the **sample size** will have to be
The more underlying **heterogeneity** (variance)	
The more **clustering** in samples	
The lower **take up**	

Cheat sheet: power calc inputs needed

- Significance and power level
  - Just use 0.05 for alpha; 0.8 for beta. Don't ask
- Main outcome(s) of interest
  - Effect size you want to be able to detect
  - Estimated standard deviation of that outcome
- For clustered RCTs
  - Estimated intracluster correlation coefficient (ICC)
  - (use command *loneway*) in stata

How can we boost power

- Focus on homogenous group (if applicable)
- High frequency data on core indicators
- Increase take up
- better quality data (its worth it…)
- Avoid clustering where possible but if its important (e.g. because of spillovers), try to maximize number of clusters