Agricultural technology & impact evaluation:
From innovation to evidence

David J. Spielman
International Food Policy Research Institute
March 2022

The historical narrative on agricultural science

- “Success stories” from advances in cultivar improvement
  - Embodied technologies: genetic traits embedded in seed
  - Seeds of change: Pro-poor, scale-neutral solutions

The new, emerging narrative

- "Success stories" that move us from improved cultivars and modern inputs to
  - Climate-smart technologies: Abiotic stress tolerant varieties
  - Disembodied technologies: Crop, water, soil, management practices
  - Knowledge-intensive technologies ≈ systems-based innovations

DSR

ZT

DTMA

FMNR

SRI

ISFM

A simple theory of change

TECHNICAL CHANGE

Adoption of new technologies and better practices

↓

ON-FARM PRODUCTIVITY GAINS

Increased yields, cost reductions, damage abatement

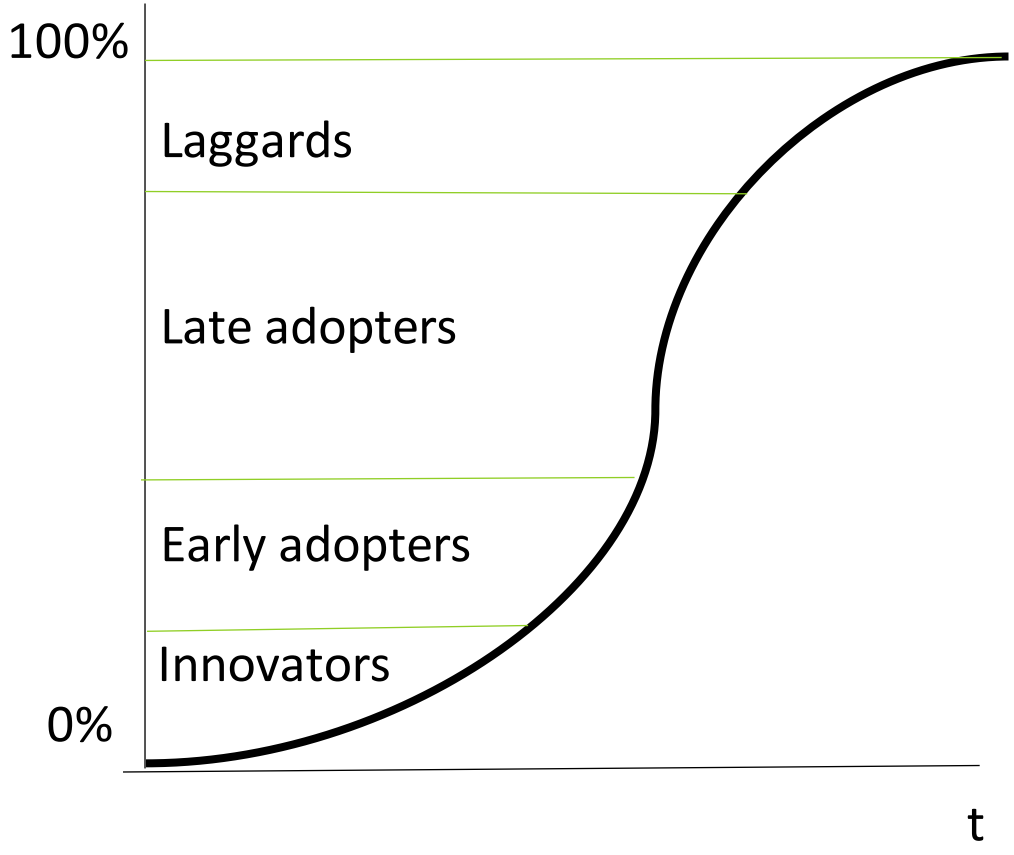
↓

IMPROVED WELFARE	GREATER SUSTAINABILITY
Better incomes, livelihoods, nutrition	Better use of scarce natural resources

Yet technology adoption is anything but simple

Constraints to adoption

- Land, soil, water, biology
- Farm size, scale, system
- Individual, household characteristics
- Access to credit, input, output markets
- Land tenure, other institutional factors
- Gross margins, net returns, variance



$$Z_i^* = \begin{cases} 1 \; if \; U_i > U_n^r \\ 0 \; if \; U_i \leq U_n^r \end{cases}$$

What *else* constrains adoption?

Lesser-known, harder-to-measure, and unobserved constraints

- Informational constraints
  - Experimentation, trialing
  - Exposure, awareness
  - Learning by doing
  - learning from others and peer effects
  - Learning by noticing
- Individual preferences
  - Risk preferences
  - Loss aversion
  - Time-inconsistent preferences and present bias
  - Aspirations

Why evaluate?

- To measure: gauge the effect caused by an intervention
- To learn: assemble evidence about what works and why
- To give feedback: improve project management and implementation
- To be accountability: accept responsibility for outcomes
- To be transparent: demonstrate a commitment to sharing information
- To design policy: credibly inform large-scale replications by government

Why do we need better designs, more rigor?

- Sample selection bias
  - Those who learn/adopt may be fundamentally different from those who don’t
  - Bias limits our ability to make wider inferences, to scale up, and to design policies
- Endogeneity
  - Reverse Causation: $X \rightarrow Y$ or $Y \rightarrow X$?
  - Simultaneity: the “Reflection Problem”
- Heterogeneity
  - *Average* effects are only so interesting
  - Measuring outcomes for specific sub-groups can paint a different picture

With a better toolkit, we can do a lot more…

- Qualitative
  - Understanding context
  - Understanding impact pathways and theories of change

- Quantitative
  - Internal validity: good identification strategies
    1. Experimental Methods: RCTs
    2. Quasi-experimental methods: D-in-D, PSM, RDD, IVs
  - External validity: generalizability

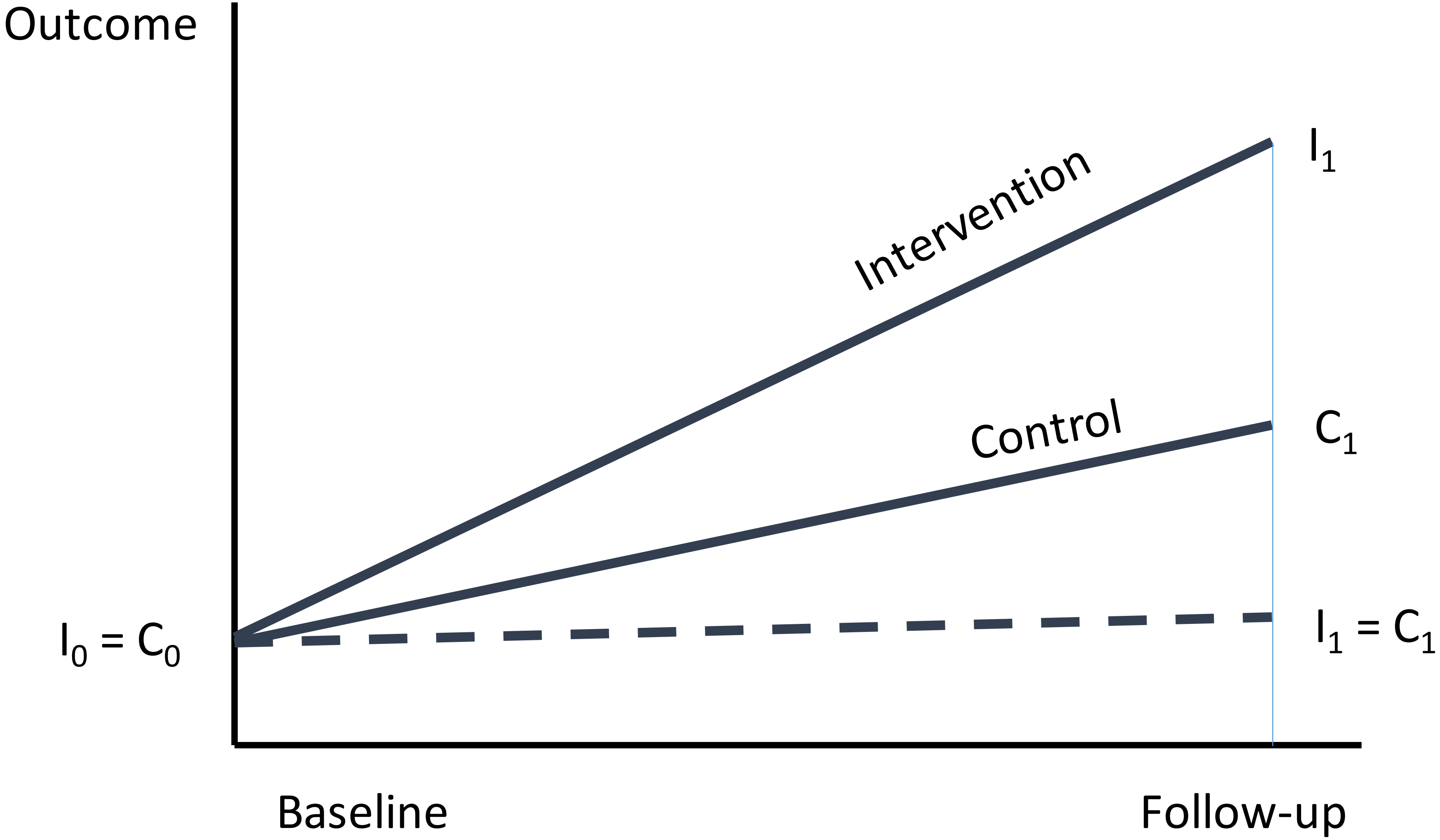
Measuring causal effects

- The causal effect of a treatment (an intervention) is the difference between outcomes with treatment (Y_1) and without it (Y_0):

$$treatment\ effect = Y_1 - Y_0$$

- Evaluation problems
  - Selection problem: individuals select into treatment if they perceive $U(Y_1) > U(Y_0)$
  - Observational problem: for the same individual, either Y_1 is observed or Y_0, but not both
  - Missing counterfactual: How much did the treated individuals benefit from the treatment compared to the situation where they were not treated?
  - Endogeneity: Does X explain Y or vice-versa?

Graphically…



Principles of impact evaluation

- We aim to make comparisons between communities, households or individuals where the ONLY difference is the treatment/intervention
- The mean and distribution of all other observable and unobservable characteristics should be the same, on average
- The first-best method of ensuring this is for the evaluator to randomly assign communities, households, or individuals to treatments and controls
- Second-best methods seek to construct these two groups artificially from available data

Why do we go through all the trouble?

- To minimize statistical bias
  - We do not want to draw inferences based on only the “best” farmers using a technology
  - We do not want to draw inferences based on only the “worst” farmers using a technology
  - We want farmers in both treatment and control groups to be similar (on average)

- To be as fair as possible
  - With a limited supply of the technology on hand, it is more equitable to randomly assign the technology to villages and farmers than to choose them *on any other basis*
  - Don’t worry: control groups can participate in the future (e.g., randomized phase-in)

- To be realistic
  - Policymakers don’t want to know how well the model farmers do with the technology; they want to know how millions of average farmers will do with the technology
  - For a scale-up of the technology, we need a sense of “what works” under real conditions for the average farmer

Experimental and quasi-experimental methods

Methods	Advantages/disadvantages
Difference of means	Difficult to isolate the true effect if treatment is not random
OLS	Does not control self-selection bias if treatment is not random
IV/2SLS	Controls selection bias, problem in finding a valid instrument
Heckman two-step estimation	Controls selection bias, does not establish a counterfactual
Difference in difference	Requires panel data to measure changes over time
Propensity score matching	Constructs a counterfactual based on matched observable characteristics
Regression discontinuity design	Estimates causal impacts using a threshold above or below which an intervention is assigned to establish a counterfactual
Randomized controlled trial	Establishes a valid counterfactual and minimizes selection bias

Treatment effects

- Average Treatment Effect: the expected causal effect of the treatment across all individuals in the population

$$ATE = [E(Y_1)|T = 1] - [E(Y_0)|T = 0]$$

- Average Treatment Effect on the Treated: the expected causal effect of the treatment for individuals in the treatment group

$$ATT = [E(Y_1 - Y_0)|T = 1]$$

- Intent to Treat Effect: the expected causal effect for individuals in the treatment group irrespective of their participation in the treatment

$$ITT = [E(Y_1)|Z = 1] - [E(Y_0)|Z = 0]$$

Exercise: Thinking like an evaluator (1)

You are an evaluator. You want to know if a new maize hybrid gives farmers higher yields than their old hybrid

Exercise 1: You estimate yield improvement by comparing farmers' yields from last season using an old hybrid against the yields they obtained this season by using the new hybrid

Question: Can you say that the difference in yields between those two season is *entirely* due to the different hybrids that were used?

Answer: NO. At least part of the differences in yields could be driven by differences in rainfall, pests, disease, or other factors that differed across these two seasons.

Exercise: Thinking like an evaluator (2)

You are an evaluator, and you still have the same question in mind

Exercise 2: You estimate yield improvement by comparing yields (in this season) on a plot where the farmer has planted the new hybrid against yields (also in this season) on the plot where she has used the old hybrid?

Question: Can you say that the difference in yields between those two plots is entirely due to the different hybrids that were used?

Answer: NO. A farmer may choose to plant the new variety on her best plot—her plot with better soil, water, etc. At least part of the differences in yields are due to differences in plot quality.

Exercise: Thinking like an evaluator (3)

You are an evaluator with that same question…

Exercise 3: Can you estimate yield improvement if you

(1) ask the farmer to divide her field in equal sub-plots
(2) randomly select half of the sub-plots and plant the new hybrid on them
(3) plant the old variety on the remaining half
(4) measure and compare yields on these plots?

Question: Can we say that the difference in yields between sub-plots planted with the new hybrid and sub-plots planted with the old hybrid is entirely due to the different hybrids that were used?

Answer: YES. Because the sub-plots were chosen randomly, the sub-plots with the new hybrid variety have the same average quality as sub-plots planted with the old hybrid.

- **But only if the random allocation was respected!!!** Only if the farmer really did plant the new and old hybrids on sub-plots that *you* selected randomly

Exercise: Thinking like an evaluator (4)

You are an evaluator with that same question, but now the Minister wants to know if the hybrid will improve yields for his constituents

Exercise 3: Can you estimate yield improvement if you identify maize farmers in his constituency, and then randomly choose half of the farmers to receive the new hybrid and half of the farmers to received the old hybrid

Question: Can we say that the difference in yields between farms planted with the new hybrid and farms planted with the old hybrid is entirely due to the different hybrids that were used?

Answer: YES. Because the farms were chosen randomly, such that for all observable and unobservable characteristics, the farms are similar – apart from the use of the new or old hybrid.

- **But only if the random allocation was respected!!!** Only if the farmer really did plant the new and old hybrids on farms that *you* selected randomly

Treatment in these exercises

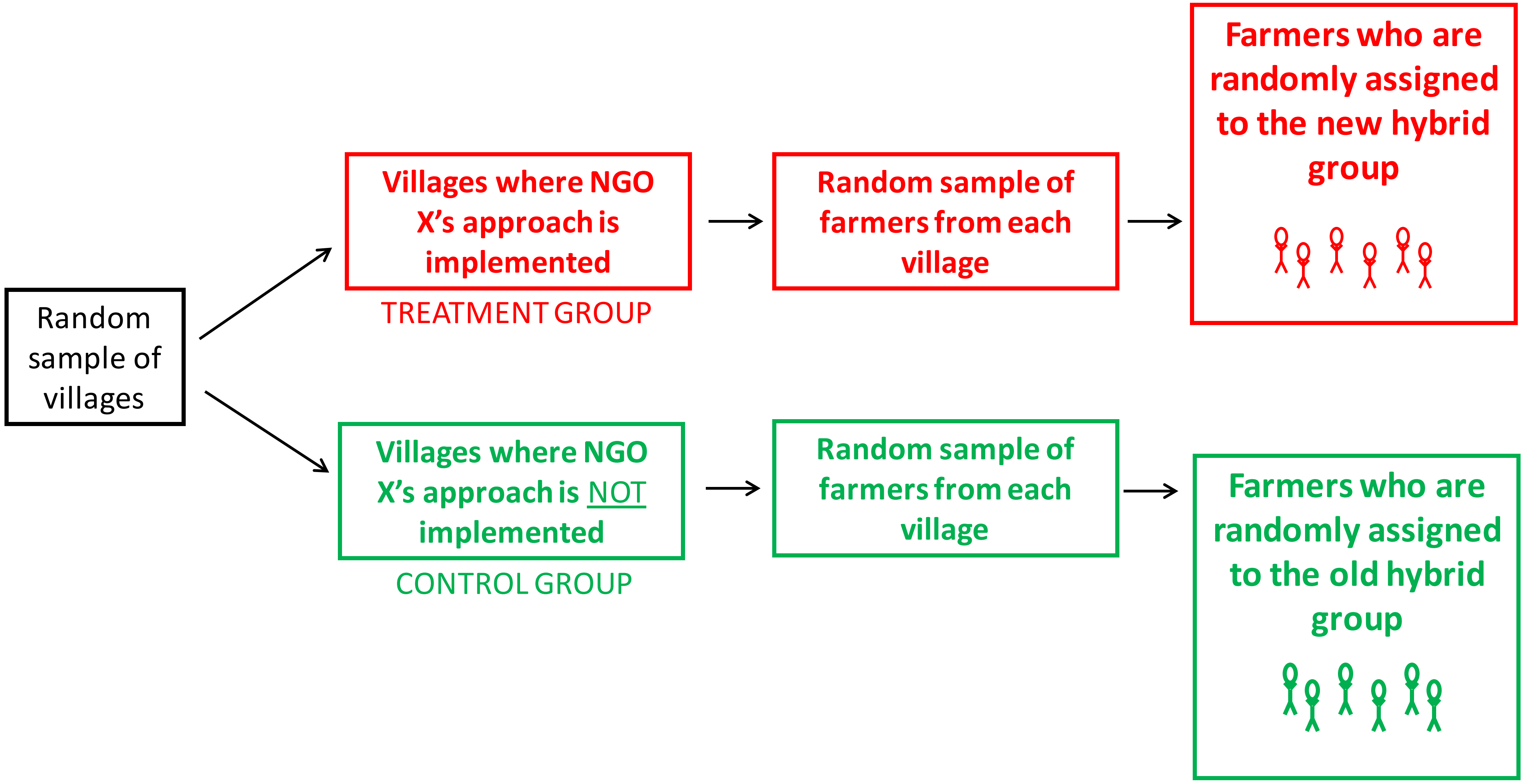
Compare

Randomly assigned individuals from locations where the treatment is implemented

to

Randomly assigned individuals from locations where the treatment is ***not*** implemented

The identification strategy in our exercise



Questions?

David J. Spielman

d.spielman@cigar.org