D4N 2023
Delivering for Nutrition in South Asia
Equity and Inclusion
November 2, 2023

Evaluating nutrition Programs at Scale:

What works?

Subir Kole, PhD
Senior Advisor – Monitoring, Evaluation & Research
Helen Keller Intl.

Agenda

1. *Context: What is an evaluation?*
2. *Commonly used evaluation methods*
3. *How do we compare outcome of interest across groups?*
4. *Avoiding pitfalls – things to consider*
5. *Wrapping up: Why evaluations are important?*

Agenda

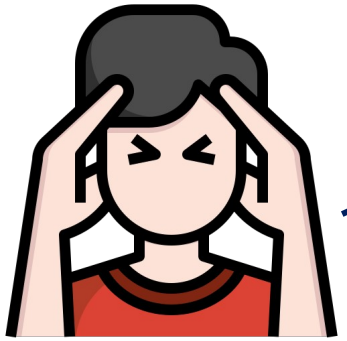
1. *Context: What is an evaluation?*
2. *Commonly used evaluation methods*
3. *How do we compare outcome of interest across groups?*
4. *Avoiding pitfalls – things to consider*
5. *Wrapping up: Why evaluations are important?*

Monitoring or Evaluation?

The Problem

The Interventions

The Result



Could be anything…!

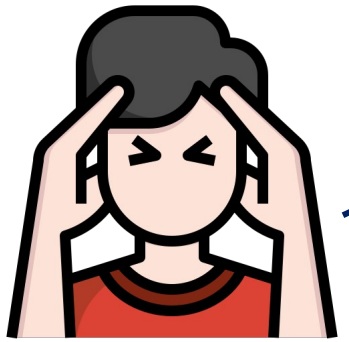
- Undernutrition
- Child mortality
- Immunization
- HIV/ TB
- NCDs….

In this simple causal framework, what/ where is “Monitoring?”

The Problem

1. Did you take the medicine (Yes/No)
2. Did you sleep for two hours after taking the medicine? (Y/N)
3. Did you switch off the lights (Yes/No)
4. Did you switch off your mobile phone or other disturbances? (Yes/No)

- Check/ Collect data
- Ongoing process…
- Result not yet achieved

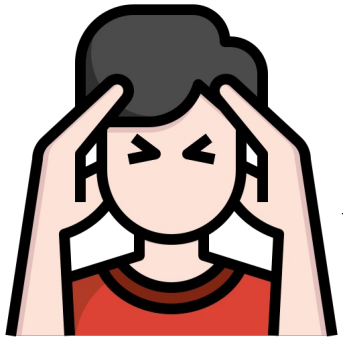


What/ where is “Evaluation” in this causal model?

The Problem

Has your migraine gone away?

1. What caused your migraine go away?
2. Is it the effect of medicine? If yes, how and to what extent?
3. Could migraine have gone away even without medicine by simply sleeping for two hours?
4. What effect does the light and sound (noise, disturbance) have in curing migraine
5. Does medicine only amplify the effect of others? If yes, to what extent?
6. Or does all four (light, sound, sleep and time) amplify the effect of medicine?
7. What if we don’t do anything?



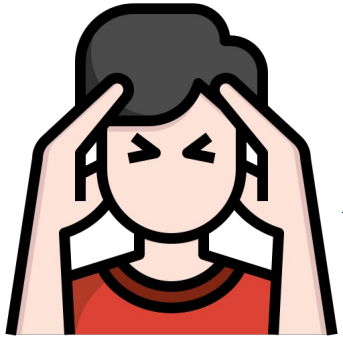
A simple headache raised multiple questions. In real world interventions, we face multiple complex questions

In evaluation, we face “Confounders”– variables that confound our understanding

The Problem

Cause/Effect relation

The Result



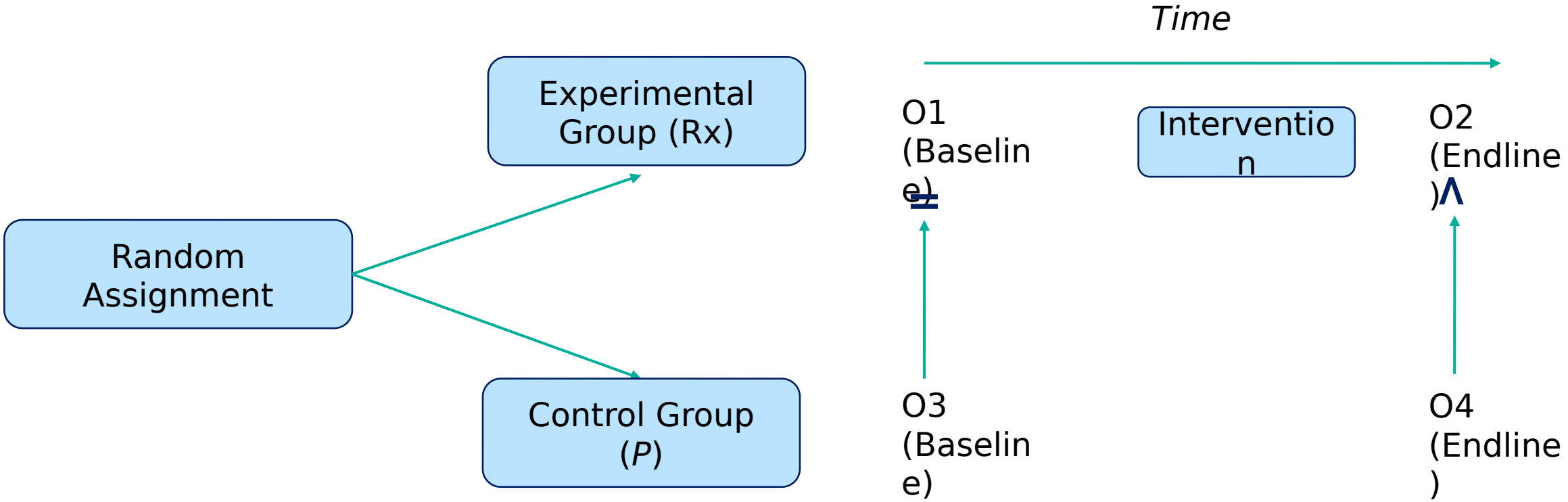
Confounders!

Example: Prevalence of lung cancer among workers in asbestos/ fireworks industry
* Some workers also smoke
What caused the cancer?

Agenda

1. *Context: What is an evaluation?*
2. *Commonly used evaluation methods*
3. *How do we compare outcome of interest across groups?*
4. *Avoiding pitfalls – things to consider*
5. *Wrapping up: Why evaluations are important?*

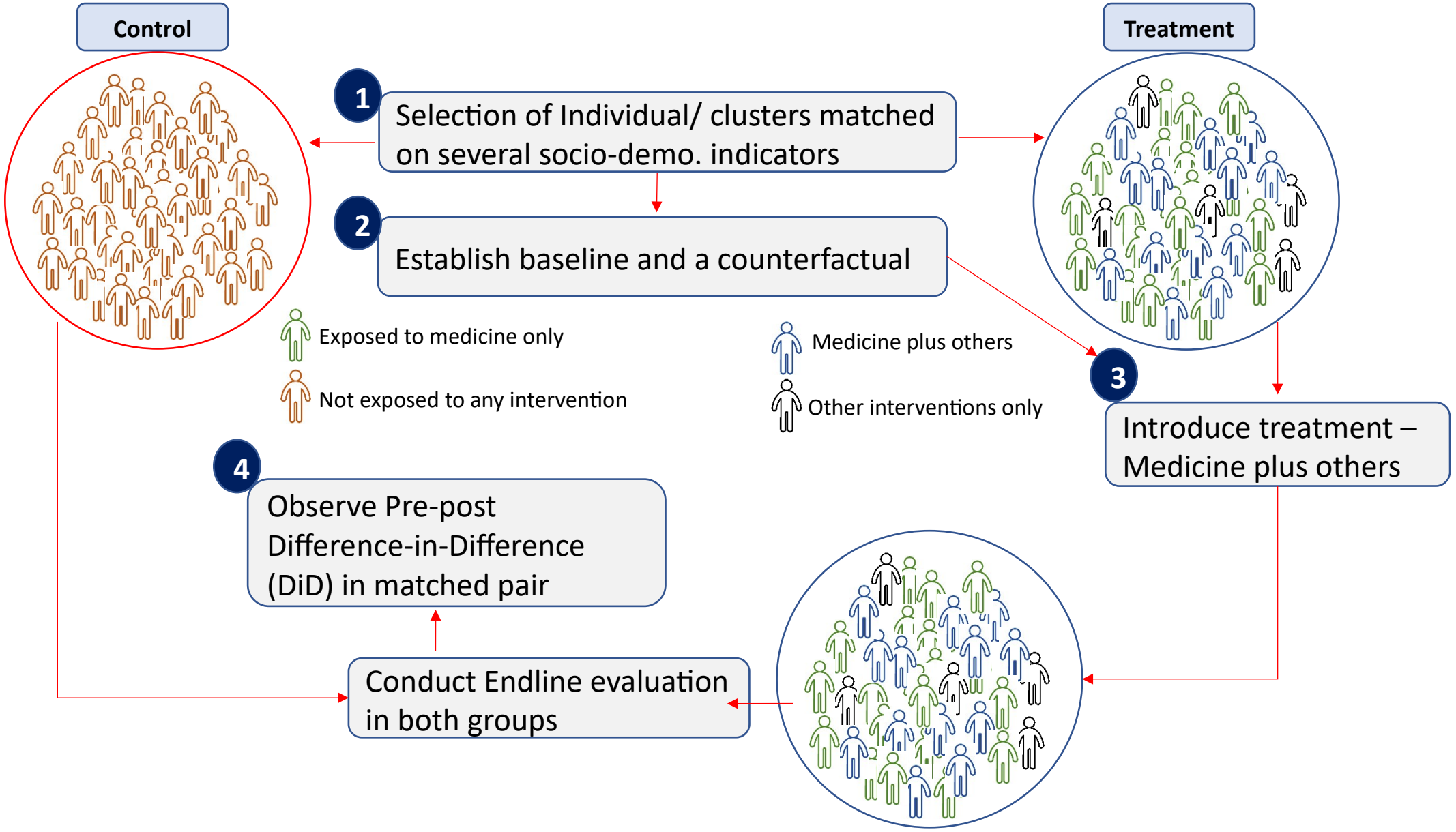
To isolate the effects of one intervention, we conduct experiments as in labs



Assumptions:

- Since Rx receives a special program, O2 would likely to be greater than O4
- Since participants are randomly assigned, O1 would be equal to O3 on key variables (age, sex, parity, education)

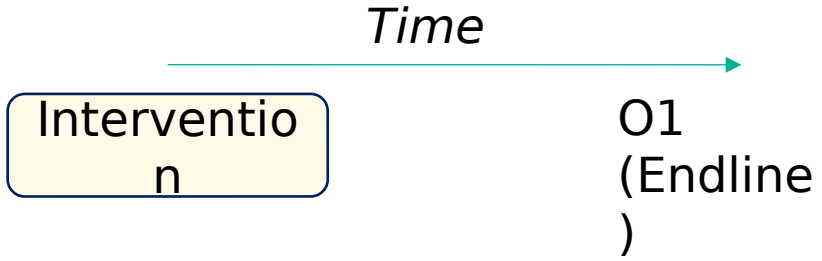
Individual/ Cluster randomized experimental design (RCT) is a gold-standard



Non-experimental Designs

1 Post-test Only design

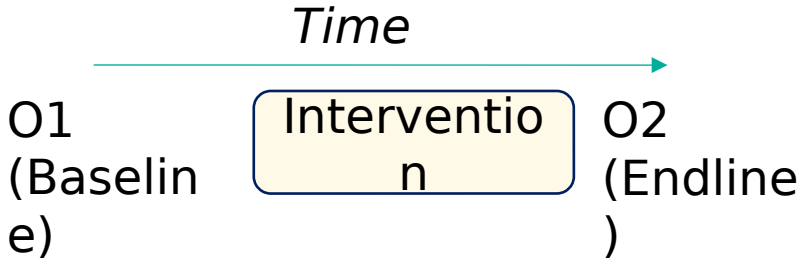
Rx group



- Validity
- Only descriptive info
- Non-comparative

2 Pretest/ Post-test design

Rx group

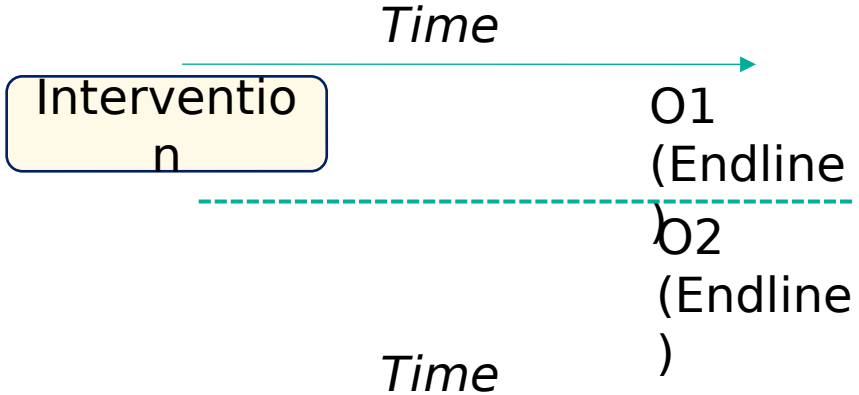


- Threats and Validity
- Only time comparison

3 Static group comparison

Rx group

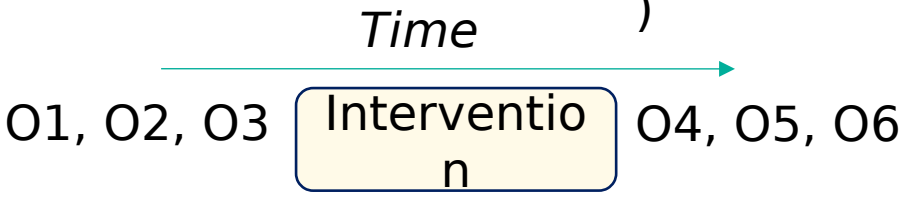
Control group



- Threats and Validity
- Non-random assignment
- Loss to follow up

4 Time Series Design

Rx group

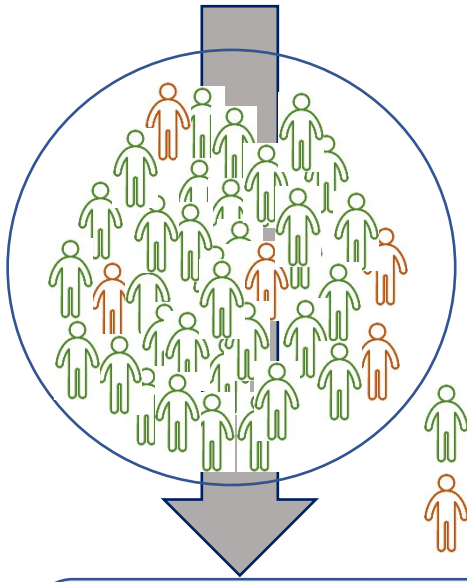


- Threats and Validity
- Only time comparison

In the absence of experimental design, we choose the least biased methods

Simple cross sectional, pre-post evaluation design

Cross-sectional



Exposed to medicines only

Not exposed to any intervention

1 Conduct a randomized cross sectional assessment of practices to treat migraine

2 Establish baseline and a counterfactual

3 Introduce treatment – Disprin plus other ways of Rx (Conf)

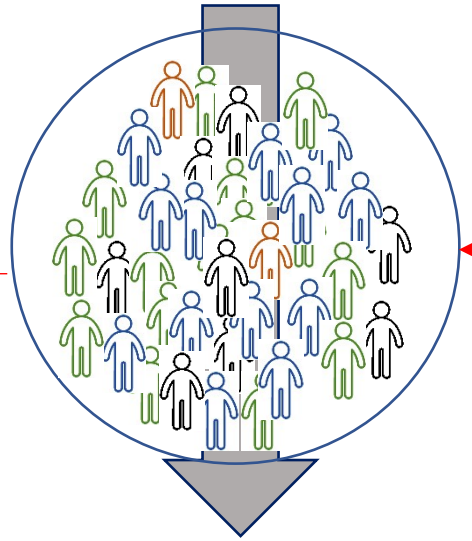
Medicine plus others

Other interventions only

Conduct Endline evaluation

- Cross sectional – randomized
- Sample size powered to detect population level changes

Cross-sectional



4 Observe the difference from baseline

Confounders – in the same geography, possibility of contamination due to mixing between groups

Selecting a specific study design is a careful consideration of several factors

Ethical Issues

- *Human subjects research*
- *Violation of people's rights/ dignity*
- *Lure/ temptation/ deception*

Practical or Administrative Issues

- *Funds, time, personnel, equipment*

Technical issues

- *Unanticipated events – closure, floods, earthquake, riots, elections*

Whenever Possible

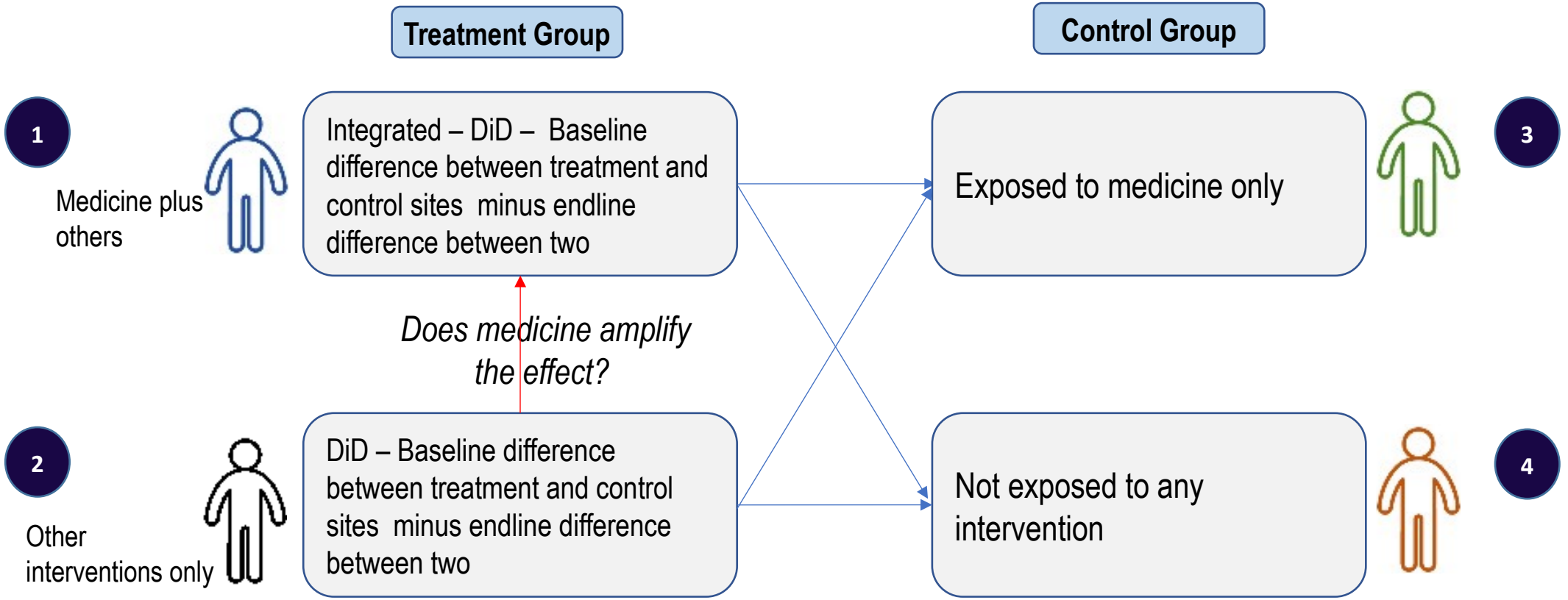
1. Adopt a random design (for assignment/ survey)
2. If 1 not possible, Choose a nearly equivalent group
3. When 1 and 2 are not possible, use a time series design
4. If 1, 2, and 3 not possible, use a pre-post design
5. If 4 is not possible, use a multivariate analytic approach
6. Keep in mind threats to validity and reliability of results

Agenda

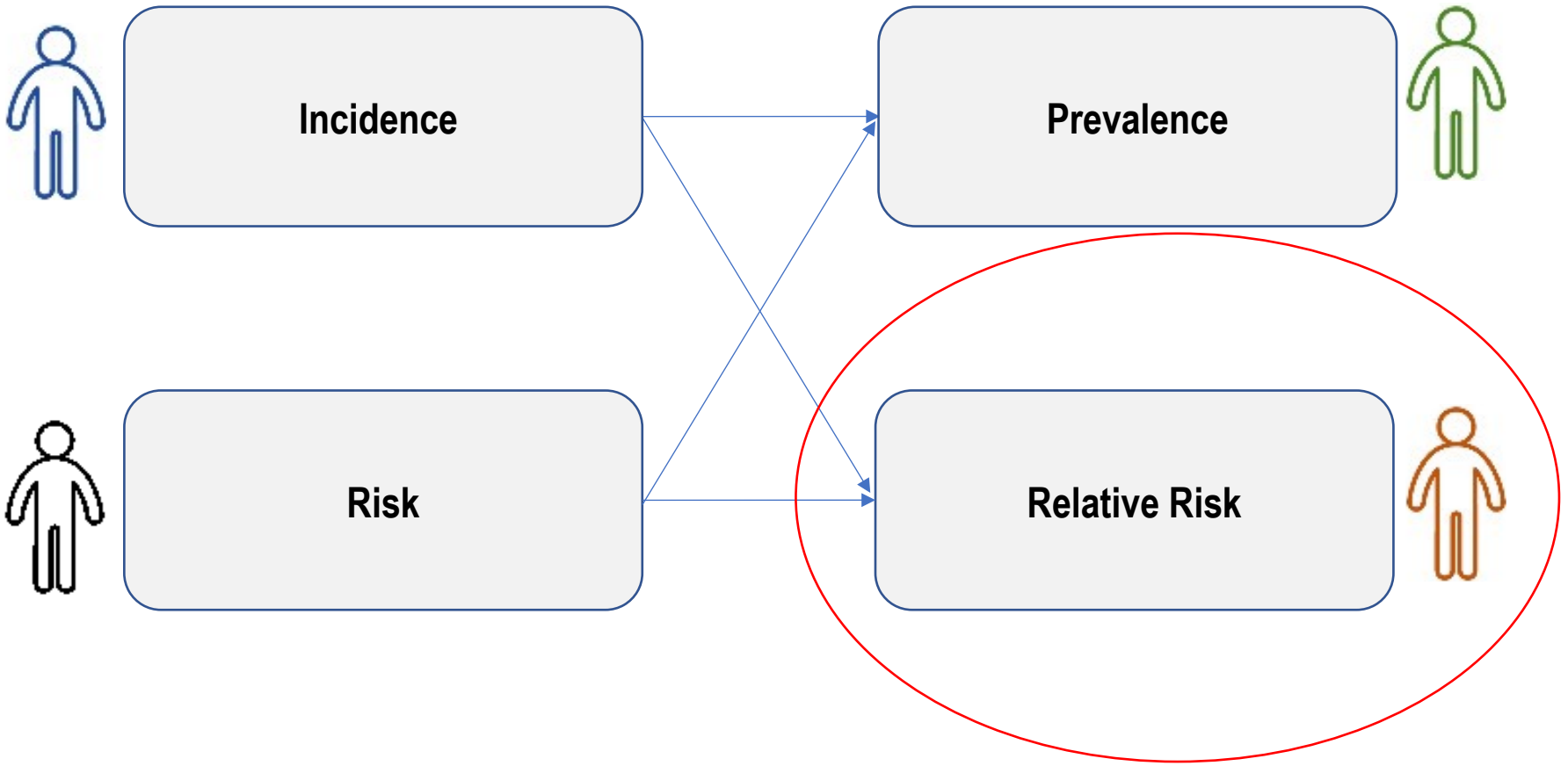
1. *Context: What is an evaluation?*
2. *Commonly used evaluation methods*
3. ***How do we compare outcome of interest across groups?***
4. *Avoiding pitfalls – things to consider*
5. *Wrapping up: Why evaluations are important?*

How do we compare outcome of interest across groups?

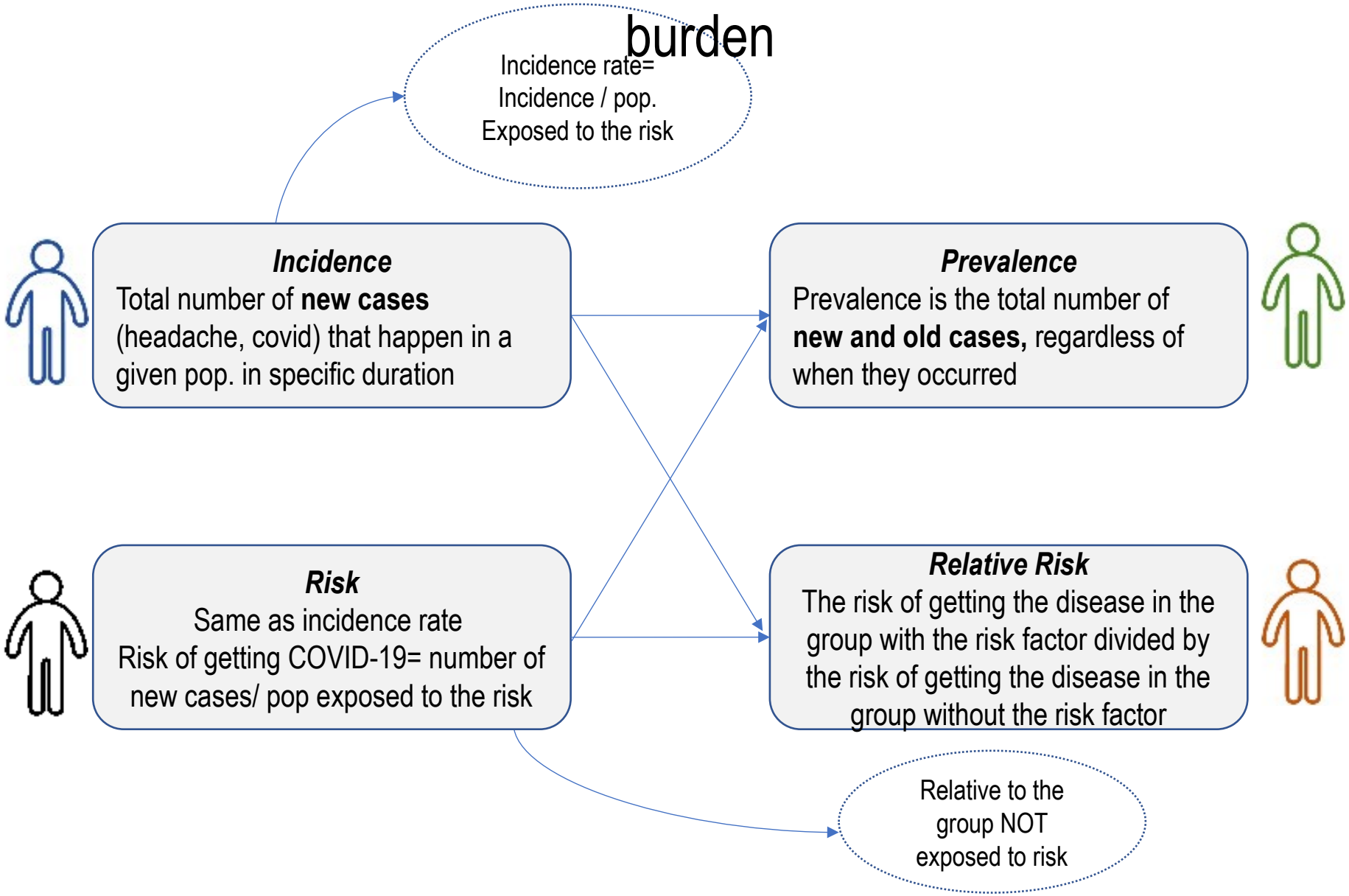
From the headache example, we observed there are four groups:



The classic 2x2 table -- basic concepts of incidence, prevalence, and risk



Incidence and prevalence are not same. Incidence= occurrence / Prevalence= burden



Understanding RR through a 2x2 table – Not using mask and covid

Risk factor present *(Not using mask)*	**Problem Present** *(COVID-19)*		**Total**
	Yes	No	
Yes	A (Subjects with the risk factor who have the problem)	B (Subjects with the risk factor who do not have the problem)	**A+B** (All subjects who have the risk factor)
No	C Subjects **without** the risk factor who have the problem	D Subjects **without** the risk factor who do not have the problem	**C+D** All subjects without the risk factor
Total	**A+C**	**B+D**	**A+B+C+D**

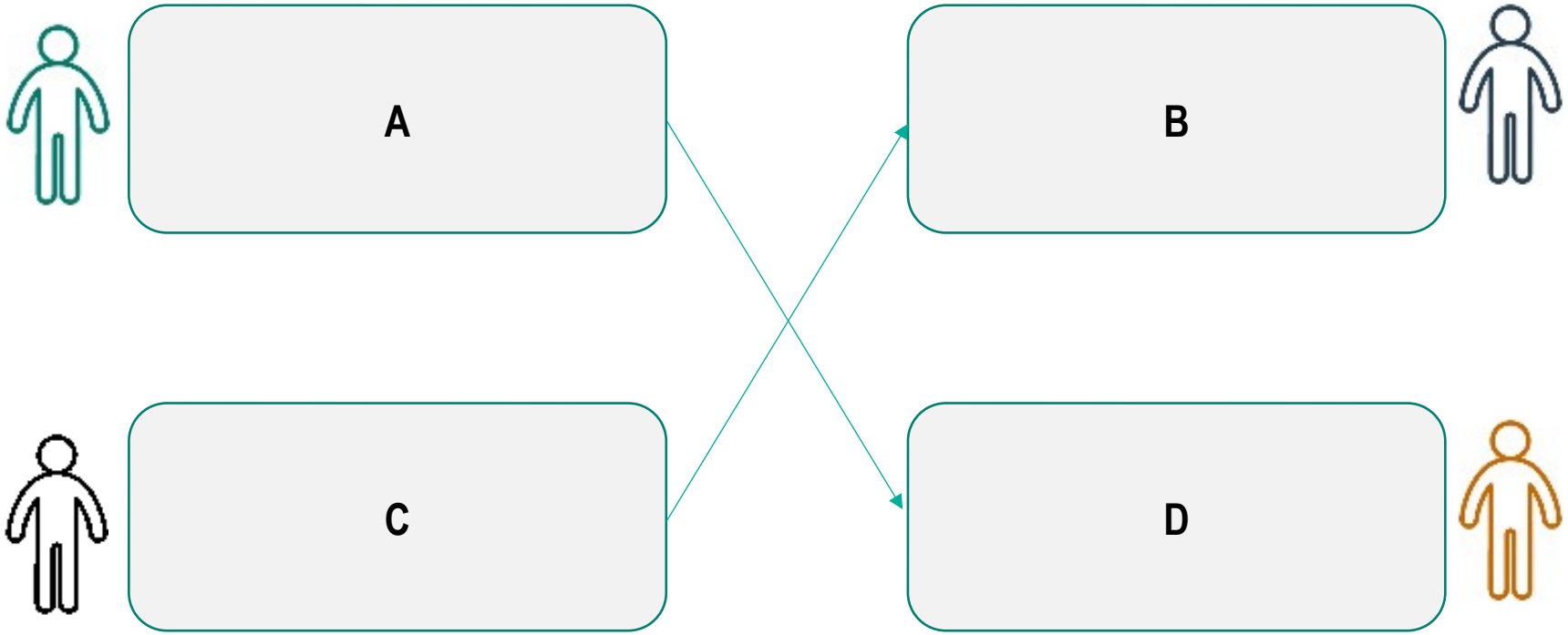
What is A/ (A+B)? → The risk of developing the problem among those with the risk factor

What is C/(C+D)? → Risk of developing the problem in those without the risk factor

What is RR? → (A/A+B) / (C/C+D)

In case/control studies, this RR is usually expressed through OR

OR measures "at the odds of" compared to a reference group



Usually expressed in times or percentages (OR= 2.54)
It is the ratio of the cross product in a 2x2 table → (AD/BC)

The basic objectives of M&E is to ensure RBM

- RBM aims to improve program effectiveness and accountability by:
  - Defining realistic expected results
  - Monitoring progress towards achievement
  - Using results for program improvement
  - Reporting on performance.

How do we track these?

- Through a concept called “results chain”
  - is a causal sequence that outlines relationship between objectives, inputs, outputs, outcomes and impact.
  - provides framework for the identification of indicators for M&E

How does a logical framework look like?

Strengthen vaccine supply chain

Vaccine planning, prioritization & Logistic management

Health workforce expansion & capacity building

Demand generation, risk communication

Community engagement & outreach

Operations research to monitor AESI

Lesser stock-outs and shortage of supplies

Skilled health workforce to deliver Covid-19 vaccine

Reduced vaccine hesitancy

Monitoring protocol for adverse events in place

Increased coverage of people with 1st/ last dose

Improved uptake of Covid-19 vaccine

Strong health system to prevent/ detect future threats

Reduced morbidity & mortality (Covid -19)

Reduction in new transmission

Inputs

Immediate Results (Outputs)

Intermediate Results (Outcomes)

Final Results (Impact)

Linking M&E with the Logical Framework

3 Monitoring Questions

2 Results Chain

1 Logframe

Are activities being implemented on time?

Actions taken or work performed to mobilize inputs

Are finance, personnel, and materials available on time & in sufficient quantities?

Financial, human & material resources required to implement the operation

Are activities leading to the immediate outputs?

The immediate, short-term products, goods and services which result from an operation

Are outputs leading to achievement of the purpose

The improved situation that an operation is expected to contribute to if completed successfully

Measuring changes here requires a longer timeframe

The highest-level result to which an organization is intended to contribute

Activities → **Inputs** → **Outputs** → **Outcomes** → **Goal/Impact**

Summarizing what's monitoring

Monitoring is the collection of routine data that measure progress toward achieving program objectives

Used to track changes in program performance over time

Sometimes referred to as process evaluation because it focuses on the implementation process

- *How well has the program been implemented?*
- *How much does implementation vary from site to site?*
- *Is it progressing towards desired objective?*

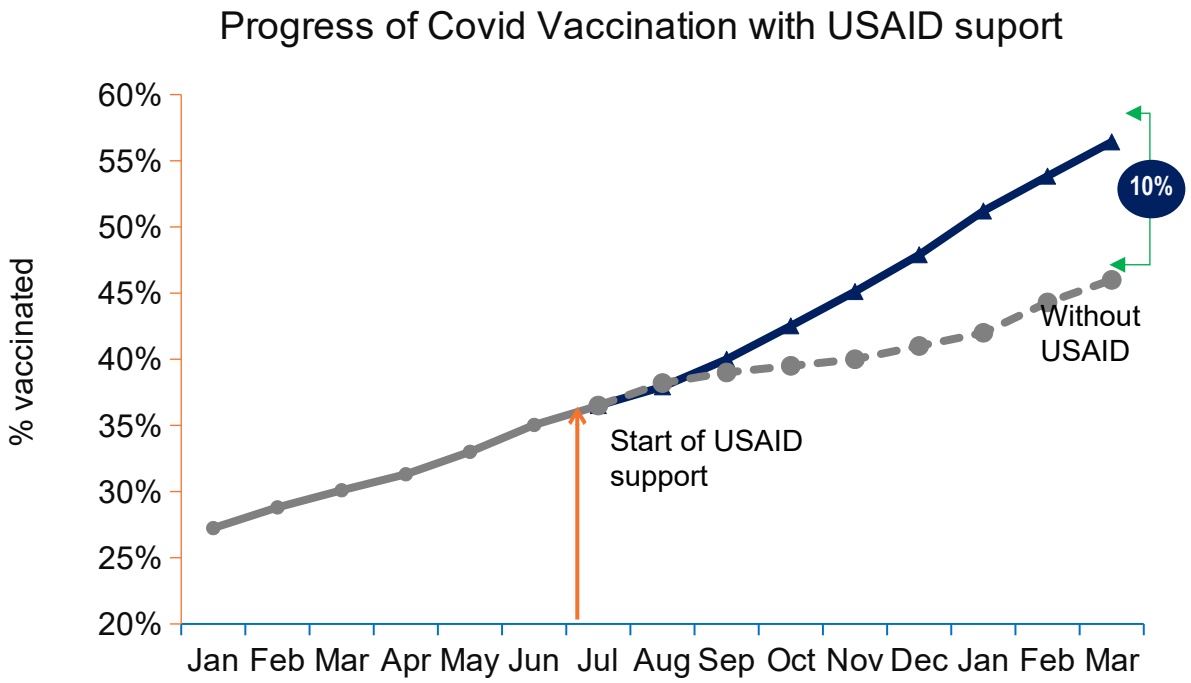
Is an ongoing, continuous process

Requires the collection of data at multiple points throughout the program cycle, including at the beginning, to provide a baseline

Involves to *counting*, *tracking*, and *collecting*

Can be used to determine if activities need adjustment during the intervention to improve desired outcomes

Summarizing evaluation…



- Evaluation measures how well the program activities have met expected objectives
- The extent to which changes in outcomes can be attributed to the intervention.
- The difference in outcome between having or not having the intervention is called “impact”

Evaluations require:

- Data collection at the start of a program (to provide a baseline) and again at the end
- A control or comparison group to measure whether the changes in outcomes can be attributed to the program
- A well-planned study design

Agenda

1. *Context: What is an evaluation?*
2. *Commonly used evaluation methods*
3. *How do we compare outcome of interest across groups?*
4. *Avoiding pitfalls – things to consider*
5. *Wrapping up: Why evaluations are important?*

Operating at scale has inherent challenges

Pitfalls	***Things to consider***
Scale-up/ Conversion of Control into Rx	Adapting/ tweaking the evaluation design. Find another matching control
Contamination/ Free mixing between groups	Replace control (??) Adjustment?
Loss of statistical power/ validity and reliability of evaluation findings	Increase sample size Or Live with it (?)
Individual tracking/ cohort monitoring Vs double/triple counting	Use a subset for cohort. Double counting is OK
Staff training – (1,100 users), Infrequent supportive supervision, coaching, guidance	Increase HR (???) May be Digital-based monitoring (photo-geo-tags)
Using service data (HMIS) creates inconsistency	Use DHS/ other survey data if objective is to bring population level change

Agenda

1. *Context: What is an evaluation?*
2. *Commonly used evaluation methods*
3. *How do we compare outcome of interest across groups?*
4. *Avoiding pitfalls – things to consider*
5. *Wrapping up: Why evaluations are important?*

Wrapping up…

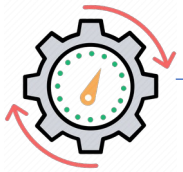
- Evaluations are important because they help tell stories of success or failure
- Generate evidence/ knowledge base on evidence-based practices for upscaling/ replication
- Fixes accountability – ensure that stakeholders are accountable for the money invested
- Promotes data driven decision making and risk management (investing in right place and deriving value for money)
- Promotes public trust and ensures transparency

Basic principles of program evaluation OECD/DAC evaluation criteria

Key evaluation questions

Relevance

- *To what extent are the objectives of the program relevant? Are the activities and outputs consistent with the overall goal? Is there adequate coverage, by activity, of the affected population? Should the program have been discontinued earlier, or should it have been extended?*



Effectiveness

- *To what extent the activities achieve its purpose / objectives? What were the major issues influencing the achievement of the objectives? Were there shared goals between different implementing agencies (coherence?) Was there effective coordination that influenced achievement of objectives?*

Efficiency

- *Were activities achieved at least cost? Were objectives achieved in a timely manner? Was the program implemented in the most efficient way compared to the alternative?*

Impact

- *What has happened as a result of the project? What real difference has the activity made to the beneficiaries? What would have happened if the program or project did not exist?*

Sustainability

- *To what extent did the program continue after external stimulus was withdrawn? What were the major factors that influenced the achievement of sustainability? Did the project address the issues of environmental, economic or social sustainability?*

Further reading

1. Robert E. Klein, Merrill S. Read et.al. (1979). Evaluating the Impact of Nutrition and Health Programs. New York: Springer.

2. Thomas J. Marchione (1984). Evaluating primary health care and nutrition programs in the context of national development. Social Science & Medicine. 19(3). pp. 225-235.

3. F. James Levinson, et.al. (2009). Monitoring and Evaluation of Nutrition Programs in Developing Countries. Nutrition Reviews 57(5). pp. 157-64.

4. OECD (2019). Better Criteria for Better Evaluation: Revised Evaluation Criteria - Definitions and Principles for Use. OECD DAC: OECD/DAC Network on Development Evaluation. https://www.oecd.org/dac/evaluation/revised-evaluation-criteria-dec-2019.pdf

Thank you!