

RealWorld Evaluation

***Designing Evaluations under Budget, Time, Data
and Political Constraints***

***WMU Evaluation Café
September 24, 2009***

by

Jim Rugh

*Note: Other PowerPoint presentations and the
summary chapter of the book are available at:*

www.RealWorldEvaluation.org

Objectives of this presentation:

1. *Very brief introduction to the seven steps of the RealWorld Evaluation approach for addressing common issues and constraints faced by evaluators such as: when the evaluator is not called in until the project is nearly completed and there was **no baseline** nor **comparison group**; or where the evaluation must be conducted with **inadequate budget** and **insufficient time**; and where there are **political pressures** and expectations for how the evaluation should be conducted and what the conclusions should say*

Objectives of this presentation:

2. *Defining what **impact evaluation** should be*
3. *Identifying and assessing various **design options** that could be used in a particular evaluation setting*
4. *Ways to **reconstruct baseline data** when the evaluation does not begin until the project is well advanced or completed.*
5. *How to minimize **threats to validity or adequacy** by utilizing appropriate combinations of quantitative and qualitative approaches (i.e. **mixed methods**) with reference to the specific context of RealWorld evaluations.*

Caveat

Note: We are focusing on project-level impact evaluations. There are, of course, many other purposes, scopes, evaluands and types of evaluations. Some of these methods may apply to them, but our examples will be based on project impact evaluations, most of them in the context of developing countries.

RealWorld Evaluation

***Designing Evaluations under Budget,
Time, Data and Political Constraints***

OVERVIEW OF THE RWE APPROACH

RealWorld Evaluation Scenarios

Scenario 1: Evaluator(s) not brought in until near end of the life of the project

For political, technical or budget reasons:

- There was no life-of-project evaluation plan
- There was no baseline survey
- Project implementers did not collect adequate data on project participants at the beginning nor during the life of the project
- It is difficult to collect data on comparable groups to get a counterfactual perspective.

A real (and recent) RealWorld Evaluation example

RFP announcement:

The aim of the 40 month project is to contribute to a sustainable improvement in the quality of rural livelihoods for up to 180,000 host and resettled returnees in the conflict affected populations of Sudan's Upper Nile and a sustainable re-integration of returnees through a community-driven recovery and rehabilitation programme [by] building capacity, improving livelihoods and increasing basic services.

A real (and recent) RealWorld Evaluation example

RFP announcement:

The evaluation should go beyond merely rating achievements against expected outputs by delving into assessing the current humanitarian and development status vis-à-vis status at the project onset so as to identify gaps that still exist and make recommendations for meeting the same.

A real (and recent) RealWorld Evaluation example

Purpose:

Assess the project impact both in quantitative and qualitative terms including the extent to which capacity building empowered LGAs and communities to identify, prioritize and meet their needs; the foundation set by livelihoods interventions in reducing food insecurity and spurring economic growth; and, the project's contribution in increasing access to and quality of basic services.

A real (and recent) RealWorld Evaluation example

Some of the goals of this evaluation:

Assess effectiveness and efficiency of the NUNRRP Consortium in light of implementation challenges.

Availability and adequacy M&E systems and the incorporation of feedback into project design and implementation.

The project's prospect for sustainability of the outputs / services / goods produced (financial viability, assets / equipment required etc)

A real (and recent) RealWorld Evaluation example

Some of the goals of this evaluation :

How cross-cutting issues such as environment and gender were mainstreamed (including involvement and participation of women in the project's decision making structures).

Assessment to the extent possible of the level of communities and local government authority involvement in the conceptualization, implementation and monitoring of the project activities.

A real (and recent) RealWorld Evaluation example

Some of the goals of this evaluation :

The extent to which the achieved outputs/outcomes meet the project overall goal (provision of peace-dividends to 180,000 target beneficiaries).

A real (and recent) RealWorld Evaluation example

Proposed approach/methodologies include:

Document review; internal and external M&E
and consultancy reports; project reports etc.

Interviews with key stakeholders

Field visits

Focus group discussions with beneficiaries and
project staff.

A real (and recent) RealWorld Evaluation example

Proposed timeline:

This evaluation is anticipated to take approximately one month, from 12 September to 6 October. [*The announcement was sent out 17 August.*]

It is the evaluator's ultimate responsibility to follow through and ensure that all relevant parties are interviewed and relevant project sites visited.

A real (and recent) RealWorld Evaluation example

Security concerns:

[The client INGO] takes no responsibility for the safety of the evaluator and / or his property during this exercise. [*Remember, this is in the Sudan.*]

A real (and recent) RealWorld Evaluation example

Qualifications:

At least an undergraduate degree and over 5 years experience in both emergency and development fields.

Experience working / or evaluating both emergency and developmental projects in sub-Saharan Africa.

Experience implementing / evaluation capacity building, livelihoods and basic services projects.

A real (and recent) RealWorld Evaluation example

Qualifications:

And be available to travel to the Sudan in three weeks, and spend a month there.

How many of you applied for that challenging evaluation consultancy assignment?

Hey, that's the RealWorld of Evaluation!!!

RealWorld Evaluation Scenarios

Scenario 2: The evaluation team is called in early in the life of the project

But for budget, political or methodological reasons:

- The 'baseline' was a needs assessment, not comparable to eventual evaluation
- It was not possible to collect baseline data on a comparison group

Reality Check – Real-World Challenges to Evaluation

- All too often, project designers do not think evaluatively – evaluation not designed until the end
- There was no baseline – at least not one with data comparable to evaluation
- There was/can be no control/comparison group.
- Limited time and resources for evaluation
- Clients have prior expectations for what the evaluation findings will say
- Many stakeholders do not understand evaluation; distrust the process; or even see it as a threat (dislike of being judged)

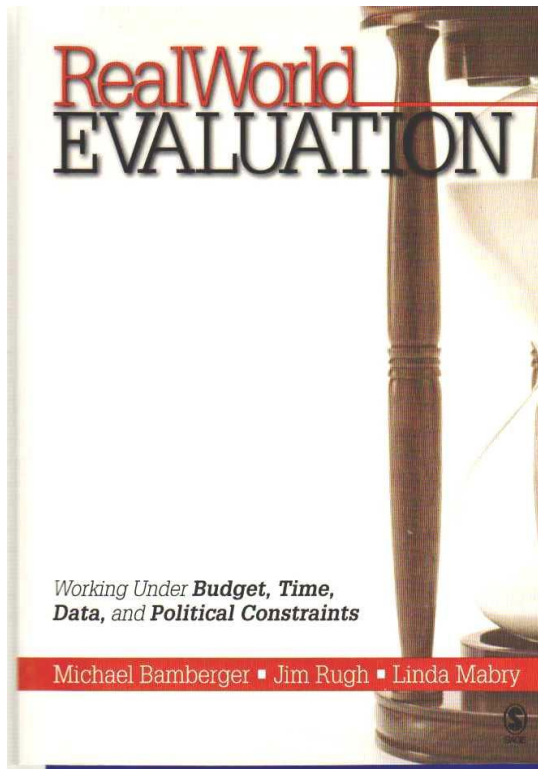
RealWorld Evaluation Quality Control Goals

- Achieve maximum possible evaluation rigor within the limitations of a given context
- Identify and control for methodological weaknesses in the evaluation design
- Negotiate with clients trade-offs between desired rigor and available resources
- Presentation of findings must acknowledge methodological weaknesses and how they affect generalization to broader populations

The Need for the RealWorld Evaluation Approach

- *As a result of these kinds of constraints, many of the basic principles of rigorous impact evaluation design (comparable pre-test-post test design, control group, adequate instrument development and testing, random sample selection, control for researcher bias, thorough documentation of the evaluation methodology etc.) are often sacrificed.*

The RealWorld Evaluation Approach



An integrated approach to ensure acceptable standards of methodological rigor while operating under RealWorld budget, time, data and political constraints.

For more details see www.RealWorldEvaluation.org for the summary chapter extracted from *RealWorld Evaluation* book, or how to order the book itself.


Most RealWorld Evaluation tools are not new— but we emphasize an holistic, integrated approach

- Most of the RealWorld Evaluation data collection and analysis tools will be familiar to most evaluators
- What is new is the **integrated approach** which combines a wide range of tools adapted to produce the best quality evaluation under RealWorld constraints
- And checklists to identify address threats to validity and adequacy

Who Uses RealWorld Evaluation and When?

- Two main users:
 - Internal and external evaluation practitioners
 - Clients, including donors and managers of implementing agencies
- The evaluation may start at:
 - the beginning of the project
 - after the project is fully operational
 - during or near the end of project implementation
 - after the project is finished

What are steps of the RealWorld Evaluation approach?

- There is a series of seven steps, each with checklists for identifying constraints and determining how to address them
 - These steps are summarized on the following slide.
- 

(There is a more detailed flow-chart in the book and in the extracted summary chapter.)

The Steps of the RealWorld Evaluation Approach

- Step 1: Planning and scoping the evaluation
- Step 2: Addressing budget constraints
- Step 3: Addressing time constraints
- Step 4: Addressing data constraints
- Step 5: Addressing political constraints
- Step 6: Assessing and addressing the strengths and weaknesses of the evaluation design
- Step 7: Helping clients use the evaluation



TIME FOR DISCUSSION

How do you define impact ?

The official OECE/DAC definition of impact is “*the positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.*”

I've observed two very different definitions of impact :

1. Effect directly attributable to a project's intervention
2. Fundamental and sustainable changes in human conditions (e.g. those identified in the MDGs and their indicators)

What does it take to measure indicators at each level?

Program Impact: Macro-level or combination of multiple project evaluations some time after projects have been completed

Project Impact : Population-based survey (baseline, evaluation)

Effect: b) Population-based survey (usually only during baseline and evaluation)

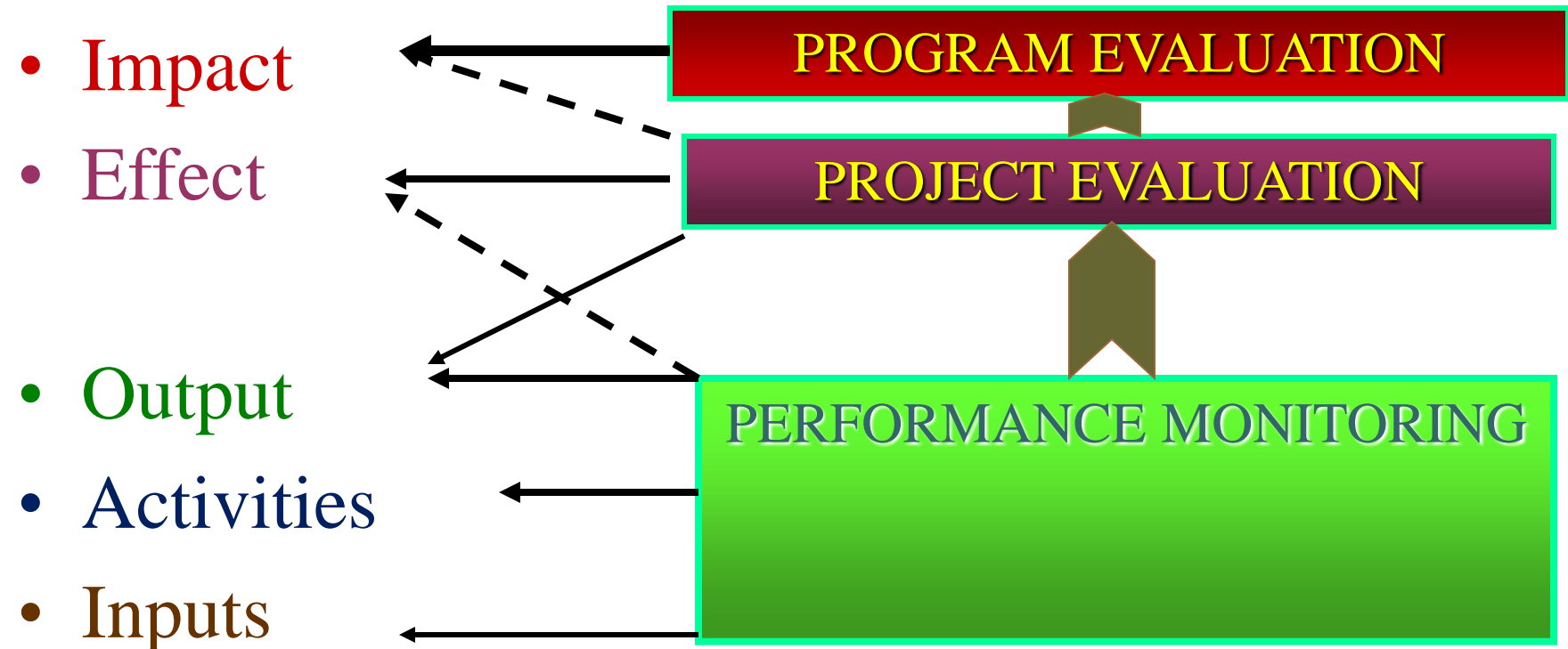
Effect: a) Follow-up survey of participants (can be done annually)

Output: Measured and reported by project staff (annually)

Activities: On-going (monitoring of interventions)

Inputs: On-going (financial accounts)

We need to recognize which evaluative process is most appropriate for measurement at various levels



Determining appropriate (and feasible) evaluation design

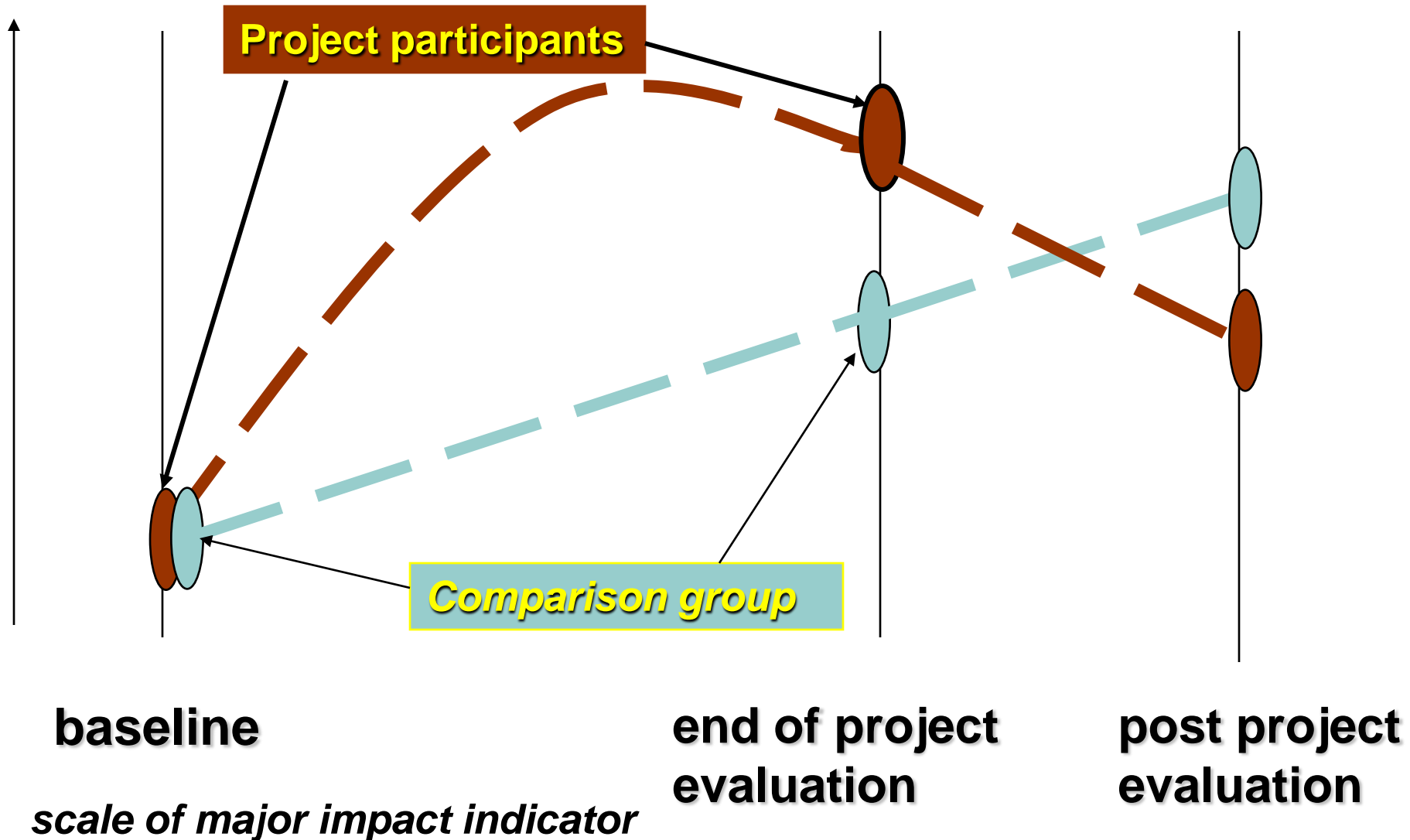
- Based on an understanding of client information needs, required level of rigor, and what is possible given the constraints, the evaluator and client need to determine what evaluation design is required and possible under the circumstances.

Let's focus for a while on evaluation design (a quick review)

- 1: Review different evaluation (experimental /quasi-experimental) research designs
- 2: The challenge of the counterfactual
- 3: Develop criteria for determining appropriate Terms of Reference (ToR) for evaluating a project, given its own (planned or un-planned) evaluation design.
- 3: A life-of-project evaluation design perspective.

An introduction to various evaluation designs

Illustrating the need for quasi-experimental longitudinal time series evaluation design



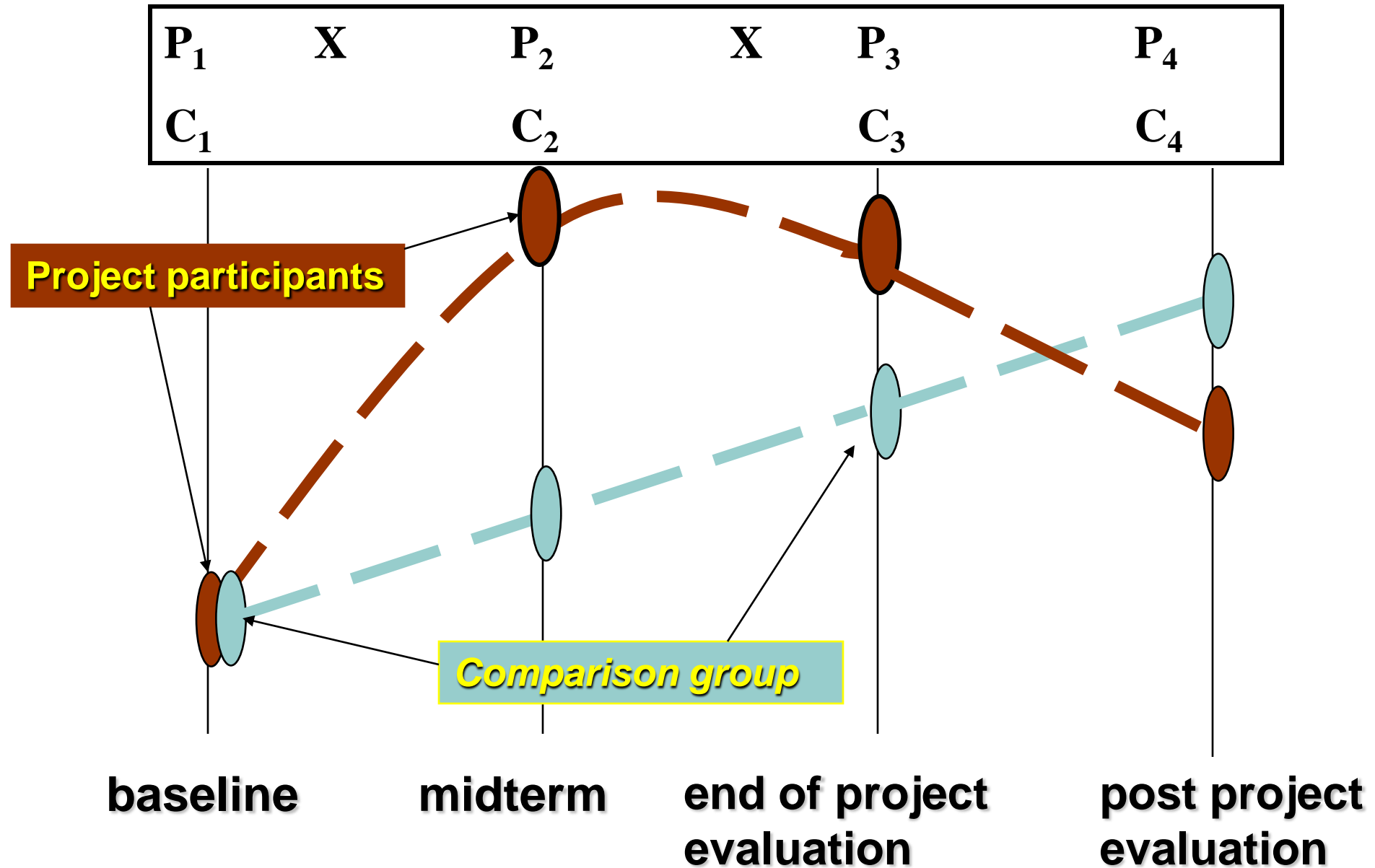
***OK, let's stop the action to
identify each of the major
types of evaluation (research)
design ...***

... one at a time, beginning with the
most rigorous design.

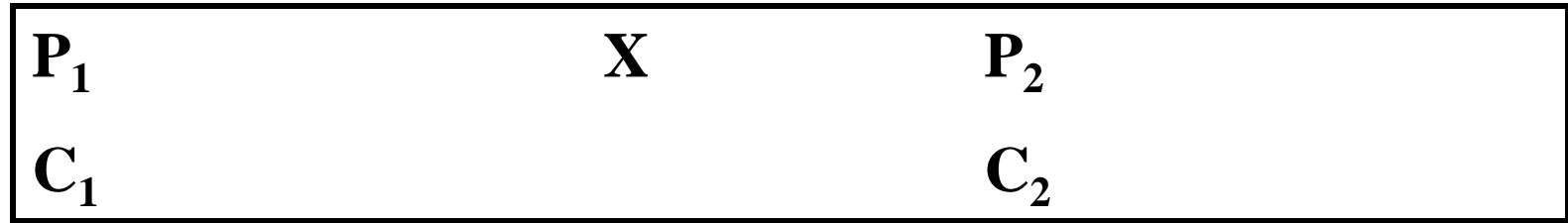
First of all: the key to the traditional symbols:

- **X** = Intervention (treatment), i.e. what the project does in a community
- **O** = Observation event (e.g. baseline, mid-term evaluation, end-of-project evaluation)
- **P** (top row): **P**roject participants
- **C** (bottom row): **C**omparison (control) group

Design #1: Longitudinal Quasi-experimental



Design #2+: Randomized Control Trial



Project participants

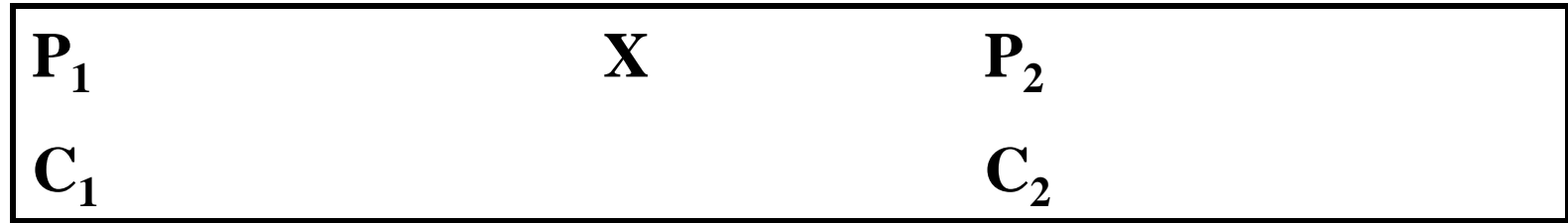
Research subjects
randomly assigned
either to project or
control group.

Control group

baseline

**end of project
evaluation**

Design #2: Quasi-experimental (pre+post, with comparison)



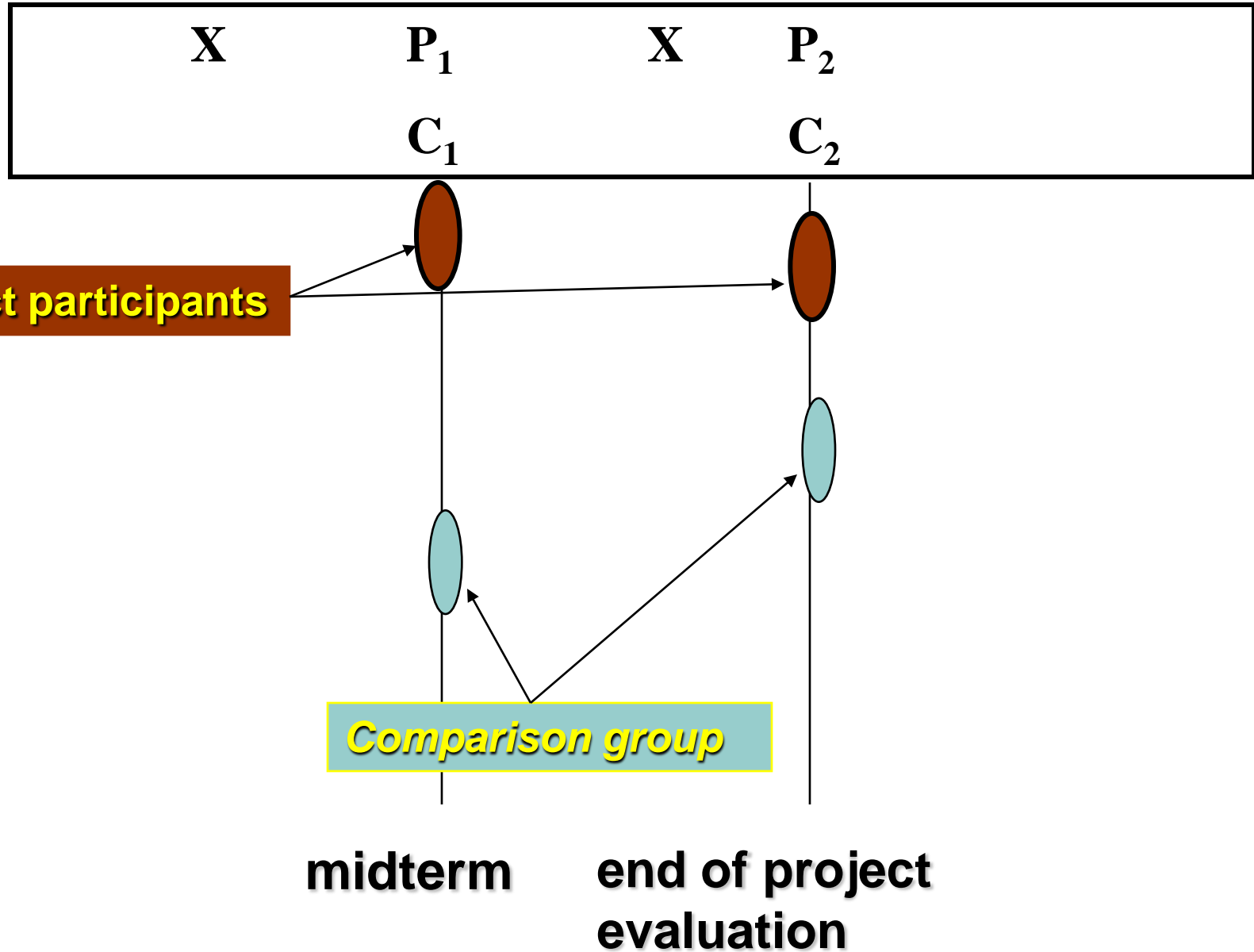
Project participants

Comparison group

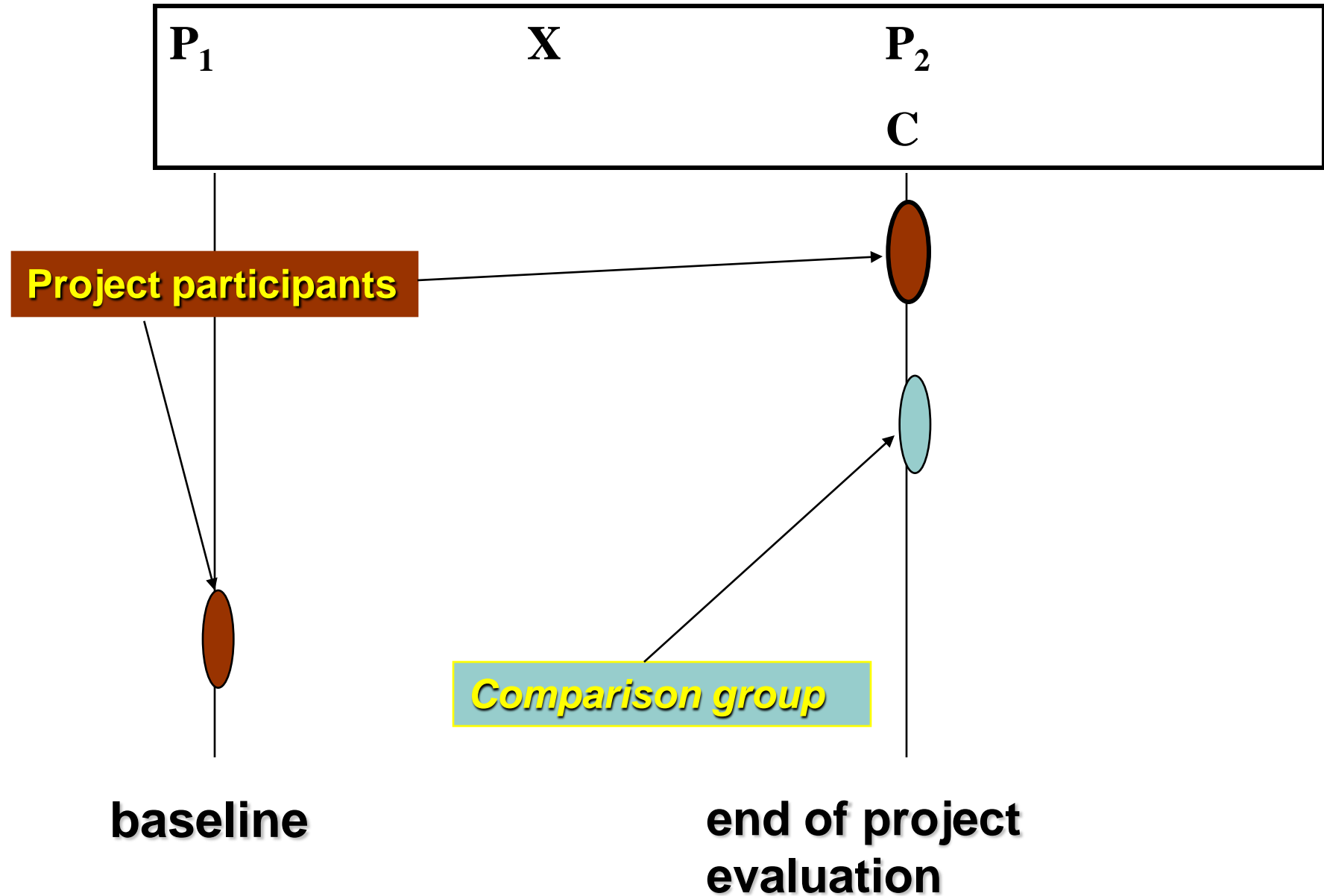
baseline

**end of project
evaluation**

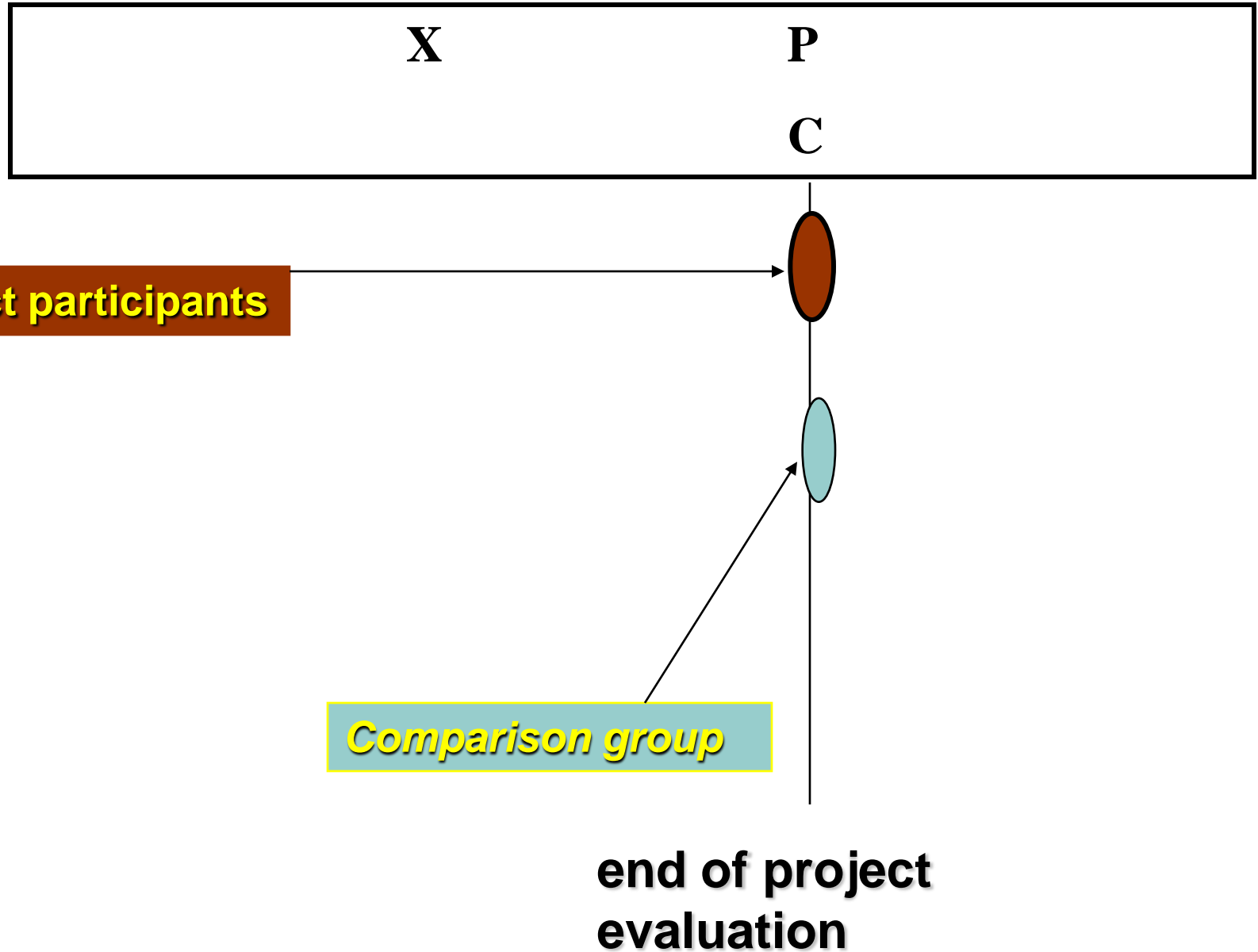
Design #3: Truncated Longitudinal



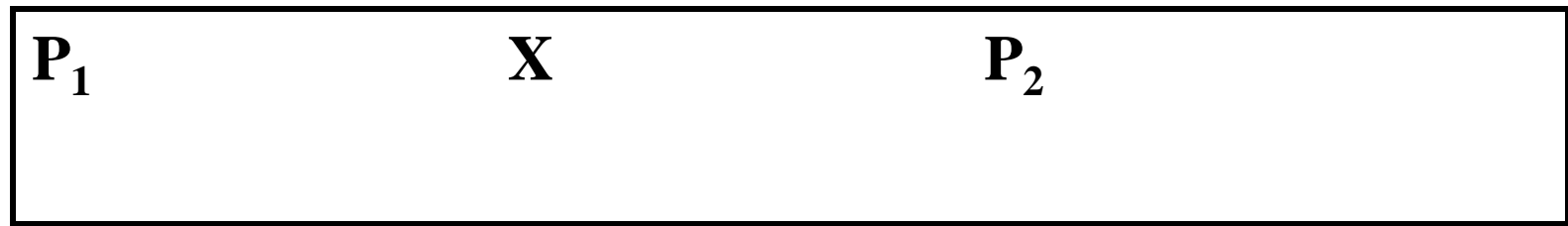
Design #4: Pre+post of project; post-only comparison



Design #5: Post-test only of project and comparison



Design #6: Pre+post of project; no comparison



Project participants



baseline



**end of project
evaluation**

Design #7: Post-test only of project participants



Project participants



**end of project
evaluation**

Attribution and counterfactuals

How do we know if the observed changes in the project participants or communities

- income, health, attitudes, school attendance etc

are due to the implementation of the project

- credit, water supply, transport vouchers, school construction etc

or to other unrelated factors?

- changes in the economy, demographic movements, other development programs etc

The Counterfactual

- What would have been the condition of the project population at the time of the evaluation if the project had not taken place?

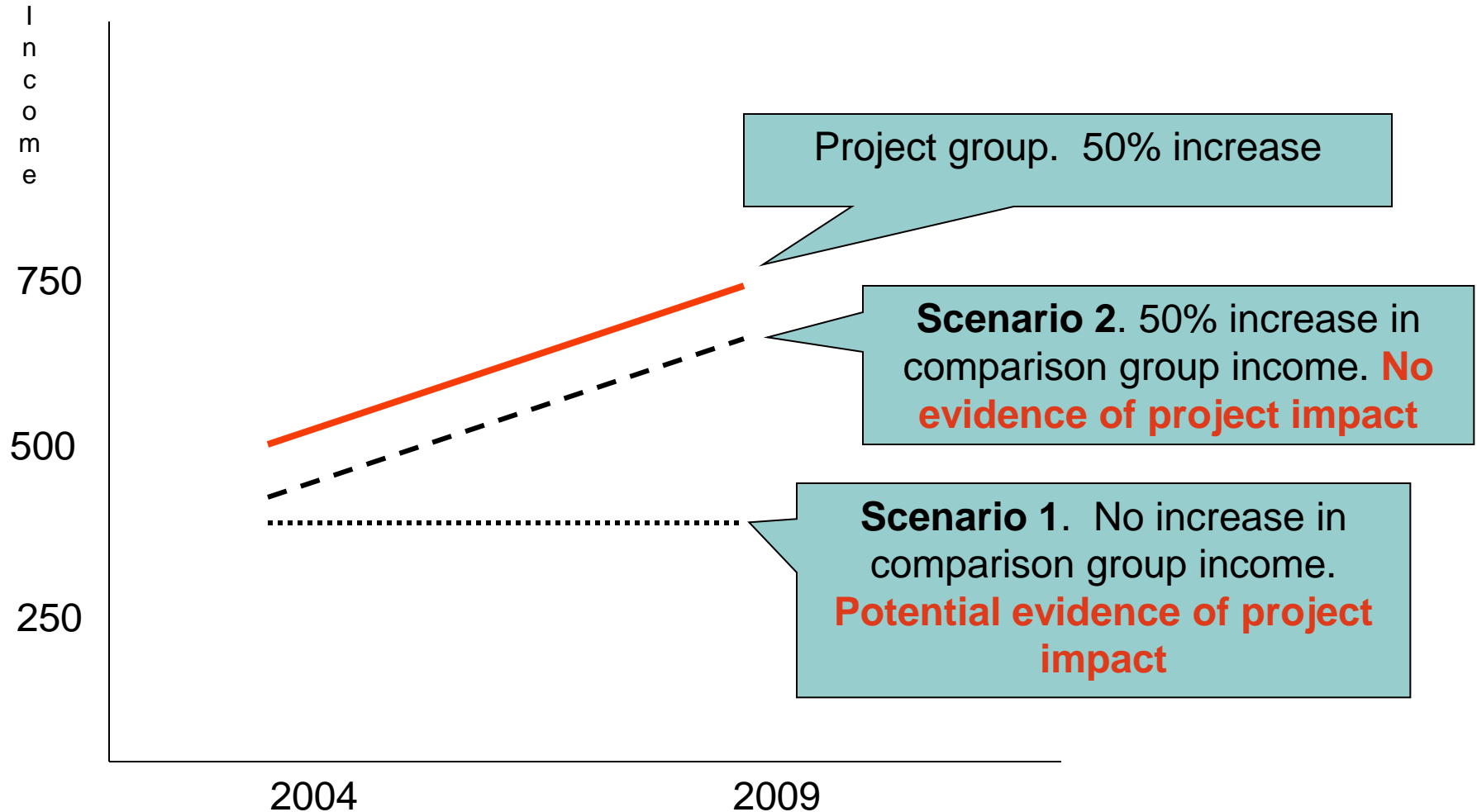
Where is the counterfactual?

After families had been living in a new housing project for 3 years, a study found average household income had increased by an 50%

Does this show that housing is an effective way to raise income?



Comparing the project with two possible comparison groups



5 main quantitative (statistical) evaluation strategies for addressing the counterfactual

Experimental designs

- I. True experimental designs
- II. Randomized selection of participants & control

Quasi-experimental designs

- III. Strong quasi-experimental designs
- IV. Weaker quasi-experimental designs

Non-experimental designs.

- V. No logically defensible counterfactual

The most rigorous statistical designs: Randomized experimental or at least strong quasi-experimental evaluation designs

	T ₁ Pre-test	T ₂ Treatment [project]	T ₃ Post-test
Project group	P ₁	X	P ₂
Control group	C ₁		C ₂

Subjects randomly assigned to the project and control groups or control group selected using statistical or judgmental matching

$$\text{Gain score [impact]} = \frac{P_2 - P_1}{C_2 - C_1}$$

Conditions of both groups are **not** controlled during the project



Control group and comparison group

- **Control group** = randomized allocation of subjects to project and non-treatment group
- **Comparison group** = separate procedure for sampling project and non-treatment groups that are as similar as possible in all aspects except the treatment (intervention)

Reference sources for randomized field trial designs

1. MIT Poverty Action Lab

www.povertyactionlab.org

2. Center for Global Development

“When will we ever learn?”

www.cgdev.org/content/publications/detail/7973

3. International Initiative for Impact Evaluation
(3ie) www.3ieimpact.org

So, is Jim saying that Randomized Control Trials are the Gold Standard and should be used in most if not all program impact evaluations?

Yes or no?

Why or why not?

If so, under what circumstances should they be used?

If not, under what circumstances would they not be appropriate?

Different lenses needed for different situations in the RealWorld

Simple	Complicated	Complex
Following a recipe	Sending a rocket to the moon	Raising a child
Recipes are tested to assure easy replication	Sending one rocket to the moon increases assurance that the next will also be a success	Raising one child provides experience but is no guarantee of success with the next
The best recipes give good results every time	There is a high degree of certainty of outcome	Uncertainty of outcome remains
Sources: Westley et al (2006) and Stacey (2007), cited in Patton 2008; also presented by Patricia Rodgers at Cairo impact conference 2009.		

Consider the RealWorld of programs to be evaluated as a giant puzzle

Experimental research (evaluation) designs, much less RCTs, are only appropriate for a few pieces of that giant puzzle.

That's why good evaluators (and those who commission evaluations) need a bigger toolbox, with a more diverse set of tools to be customized when designing evaluations that respond to different purposes and circumstances.

The limited use of strong evaluation designs

- In the realworld we estimate that
 - fewer than 5%-10% of impact evaluations use a strong quasi-experimental design
 - significantly less than 5% use randomized control trials ('pure, scientific' experimental design)

There are other methods for assessing the counterfactual

- Reliable secondary data that depicts relevant trends in the population
- Longitudinal monitoring data (if it includes non-reached population)
- Qualitative methods to obtain perspectives of key informants, participants, neighbors, etc.

Other questions to answer as you customize an evaluation Terms of Reference (ToR):

1. Who asked for the evaluation? (Who are the key stakeholders)?
2. What are the key questions to be answered?
3. Will this be a formative or summative evaluation?
4. Will there be a next phase, or other projects designed based on the findings of this evaluation?

Other questions to answer as you customize an evaluation ToR:

5. What decisions will be made in response to the findings of this evaluation?
6. What is the appropriate level of rigor?
7. What is the scope / scale of the evaluation / evaluand (thing to be evaluated)?
8. How much time will be needed / available?
9. What financial resources are needed / available?

Other questions to answer as you customize an evaluation ToR:

10. Should the evaluation rely mainly on quantitative or qualitative methods?
11. Should participatory methods be used?
12. Can / should there be a household survey?
13. Who should be interviewed?
14. Who should be involved in planning / implementing the evaluation?
15. What are the most appropriate media for communicating the findings to different stakeholder audiences?

Evaluation (research) design?

Resources available?

Key questions?

Time available?

Evaluand (what to evaluate)?

Qualitative?

Skills available?

Quantitative?

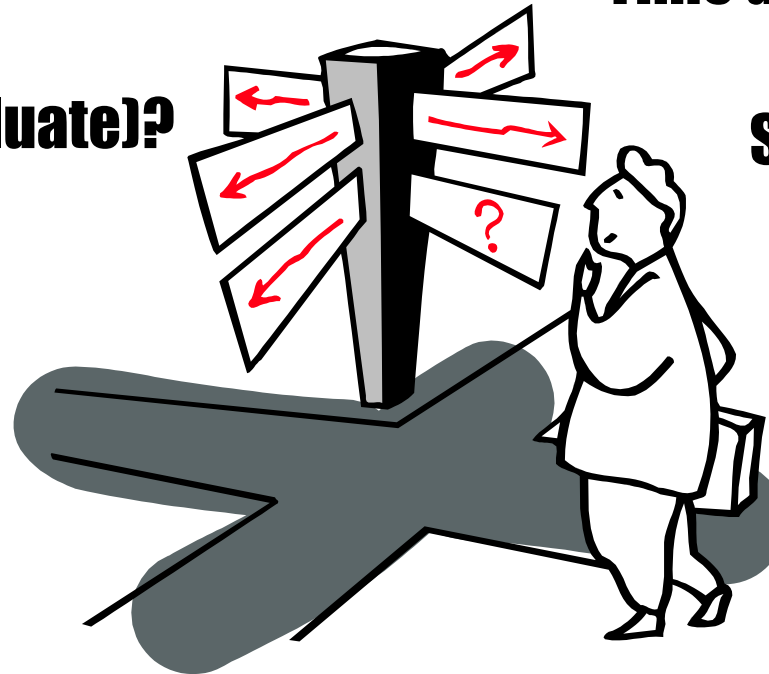
Participatory?

Scope?

Extractive?

Appropriate level of rigor?

Evaluation FOR whom?



Does this help, or just confuse things more? Who said evaluations (like life) would be easy?!!

**Before we return to
the RealWorld steps,
let's gain a
perspective on levels
of rigor, and what a
life-of-project
evaluation plan could
look like**

Different levels of rigor

depends on source of evidence; *level of confidence*; *use of information*

Objective, high precision – but requiring more time & expense

Level 5: A very thorough research project is undertaken to conduct in-depth analysis of situation; *P= +/- 1%* *Book published!*

Level 4: Good sampling and data collection methods used to gather data that is representative of target population; *P= +/- 5%* *Decision maker reads full report*

Level 3: A rapid survey is conducted on a convenient sample of participants; *P= +/- 10%* *Decision maker reads 10-page summary of report*

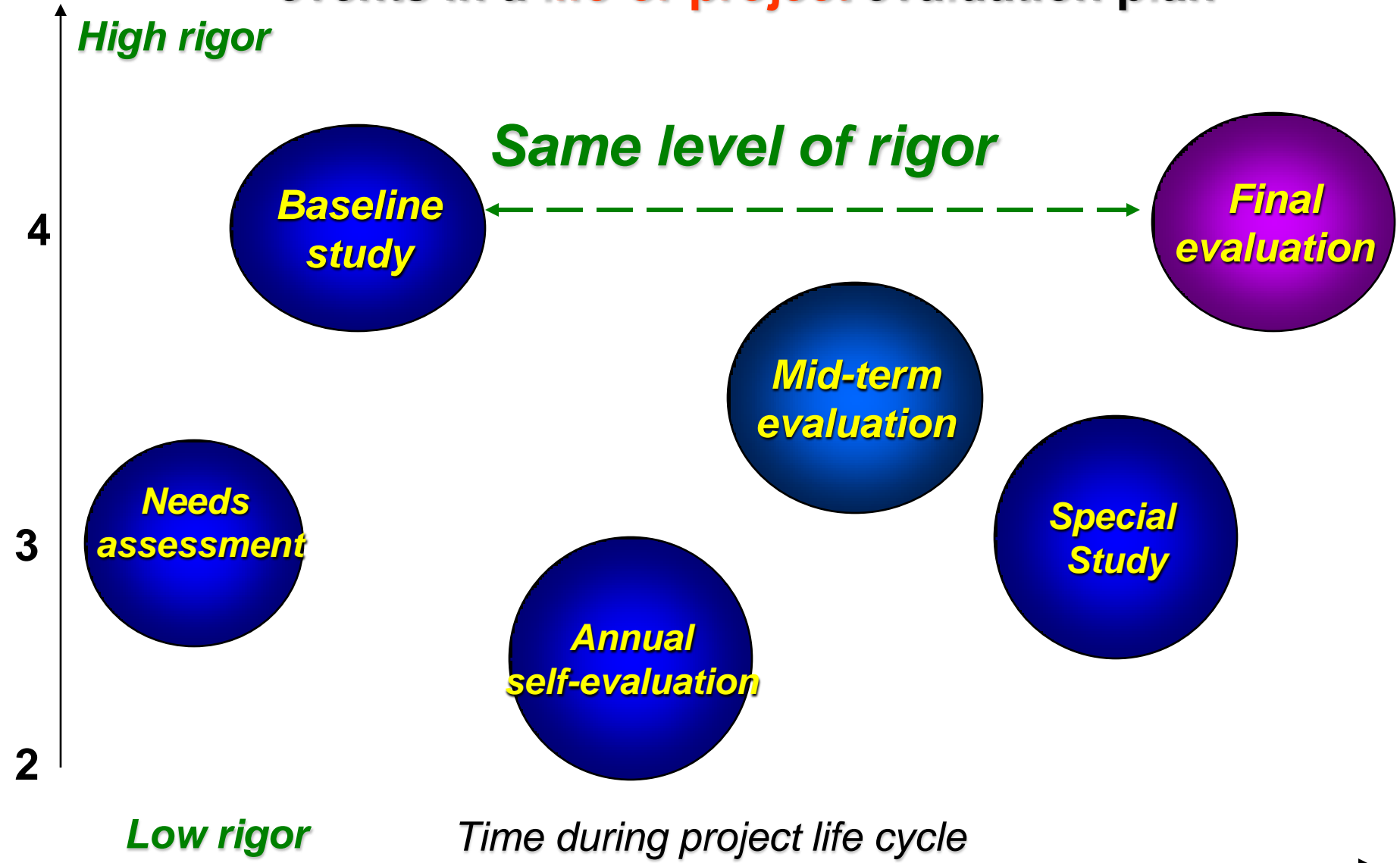
Level 2: A fairly good mix of people are asked their perspectives about project; *P= +/- 25%* *Decision maker reads at least executive summary of report*

Level 1: A few people are asked their perspectives about project; *P= +/- 40%* *Decision made in a few minutes*

Level 0: Decision-maker's impressions based on anecdotes and sound bytes heard during brief encounters (hallway gossip), mostly intuition; *Level of confidence +/- 50%*; *Decision made in a few seconds*

Quick & cheap – but subjective, sloppy

Determining appropriate **levels of precision** for events in a **life-of-project** evaluation plan





TIME FOR DISCUSSION

RealWorld Evaluation

***Designing Evaluations under Budget,
Time, Data and Political Constraints***

Steps 2 + 3

***ADDRESSING BUDGET AND
TIME CONSTRAINTS***

Step 2: Addressing budget constraints

- A. Clarifying client information needs
- B. Simplifying the evaluation design
- C. Look for reliable secondary data
- D. Review sample size
- E. Reducing costs of data collection and analysis

Look for reliable secondary sources

- Planning studies, project administrative records, government ministries, other NGOs, universities / research institutes, mass media.

Look for reliable secondary sources, cont.

Assess the relevance and reliability of sources for the evaluation with respect to:

- Coverage of the target population
- Time period
- Relevance of the information collected
- Reliability and completeness of the data
- Potential biases

Some ways to save time and money

- Depending upon the purpose and level of rigor required, some of the options might include:
 - Reducing the number of units studied (communities, families, schools)
 - Reducing the number of case studies or the duration and complexity of the cases
 - Reducing the duration or frequency of observations

Seeking ways to reduce sample size

Accepting a lower level of precision significantly reduces the required number of interviews:

- To test for a **5%** change in proportions requires a minimum sample of **1086**
- To test for a **10%** change in proportions requires a minimum sample of **270**

Reducing costs of data collection and analysis

- Use self-administered questionnaires
- Reduce length and complexity of survey instrument
- Use direct observation
- Obtain estimates from focus groups and community forums
- Key informants
- Participatory assessment methods
- Multi-methods and triangulation

Step 3: Addressing time constraints

In addition to Step 2 methods:

- Reduce time pressures on external consultants
 - Commission preparatory studies
 - Video conferences
- Hire more consultants/researchers
- Incorporate outcome indicators in project monitoring systems and documents
- Technology for data inputting/coding

Addressing time constraints

Negotiate with the client to discuss questions such as the following:

1. What information is essential and what could be dropped or reduced?
2. How much precision and detail is required for the essential information? E.g. is it necessary to have separate estimates for each geographical region or sub-group or is a population average acceptable?
3. Is it necessary to analyze all project components and services or only the most important?
4. Is it possible to obtain additional resources (money, staff, computer access, vehicles etc) to speed up the data collection and analysis process?

RealWorld Evaluation

***Designing Evaluations under Budget,
Time, Data and Political Constraints***

Step 4

**Addressing data
constraints**

Ways to reconstruct baseline conditions

- A. Secondary data
- B. Project records
- C. Recall
- D. Key informants

Ways to reconstruct baseline conditions

- E. PRA (Participatory Rapid Appraisal) and PLA (Participatory Learning and Action) and other participatory techniques such as timelines and critical incidents to help establish the chronology of important changes in the community

Assessing the utility of potential secondary data

- Reference period
- Population coverage
- Inclusion of required indicators
- Completeness
- Accuracy
- Free from bias

Examples of secondary data to reconstruct baselines

- Census
- Other surveys by government agencies
- Special studies by NGOs, donors
- University research studies
- Mass media (newspapers, radio, TV)
- External trend data that might have been monitored by implementing agency

Using recall to reconstruct baseline data

- School attendance and time/cost of travel
- Sickness/use of health facilities
- Income and expenditures
- Community/individual knowledge and skills
- Social cohesion/conflict
- Water usage/quality/cost
- Periods of stress
- Travel patterns

Limitations of recall

- Generally not reliable for precise quantitative data
- Sample selection bias
- Deliberate or unintentional distortion
- Few empirical studies (except on expenditure) to help adjust estimates

Sources of bias in recall

- Who provides the information
- Under-estimation of small and routine expenditures
- “Telescoping” of recall concerning major expenditures
- Distortion to conform to accepted behavior:
 - Intentional or unconscious
 - Romanticizing the past
 - Exaggerating (e.g. “We had nothing before this project came!”)
- Contextual factors:
 - Time intervals used in question
 - Respondents expectations of what interviewer wants to know
- Implications for the interview protocol

Improving the validity of recall

- Conduct small studies to compare recall with survey or other findings.
- Ensure all relevant groups interviewed
- Triangulation
- Link recall to important reference events
 - Elections
 - Drought/flood/tsunami/war/displacement
 - Construction of road, school etc

Key informants

- Not just officials and high status people
- Everyone can be a key informant on their own situation:
 - Single mothers
 - Factory workers
 - Users of public transport
 - Sex workers
 - Street children

Guidelines for key-informant analysis

- Triangulation greatly enhances validity and understanding
- Include informants with different experiences and perspectives
- Understand how each informant fits into the picture
- Employ multiple rounds if necessary
- Carefully manage ethical issues

PRA and related participatory techniques

- PRA (Participatory Rapid Appraisal) and PLA (Participatory Learning and Action) techniques collect data at the group or community [rather than individual] level
- Can either seek to identify consensus or identify different perspectives
- Risk of bias:
 - If only certain sectors of the community participate
 - If certain people dominate the discussion

Summary of issues in baseline reconstruction

- Variations in reliability of recall
- Memory distortion
- Secondary data not easy to use
- Secondary data incomplete or unreliable
- Key informants may distort the past

2. Ways to reconstruct comparison groups

- Judgmental matching of communities
- When there is phased introduction of project services beneficiaries entering in later phases can be used as “pipeline” comparison groups
- Internal controls when different subjects receive different combinations and levels of services

RealWorld Evaluation

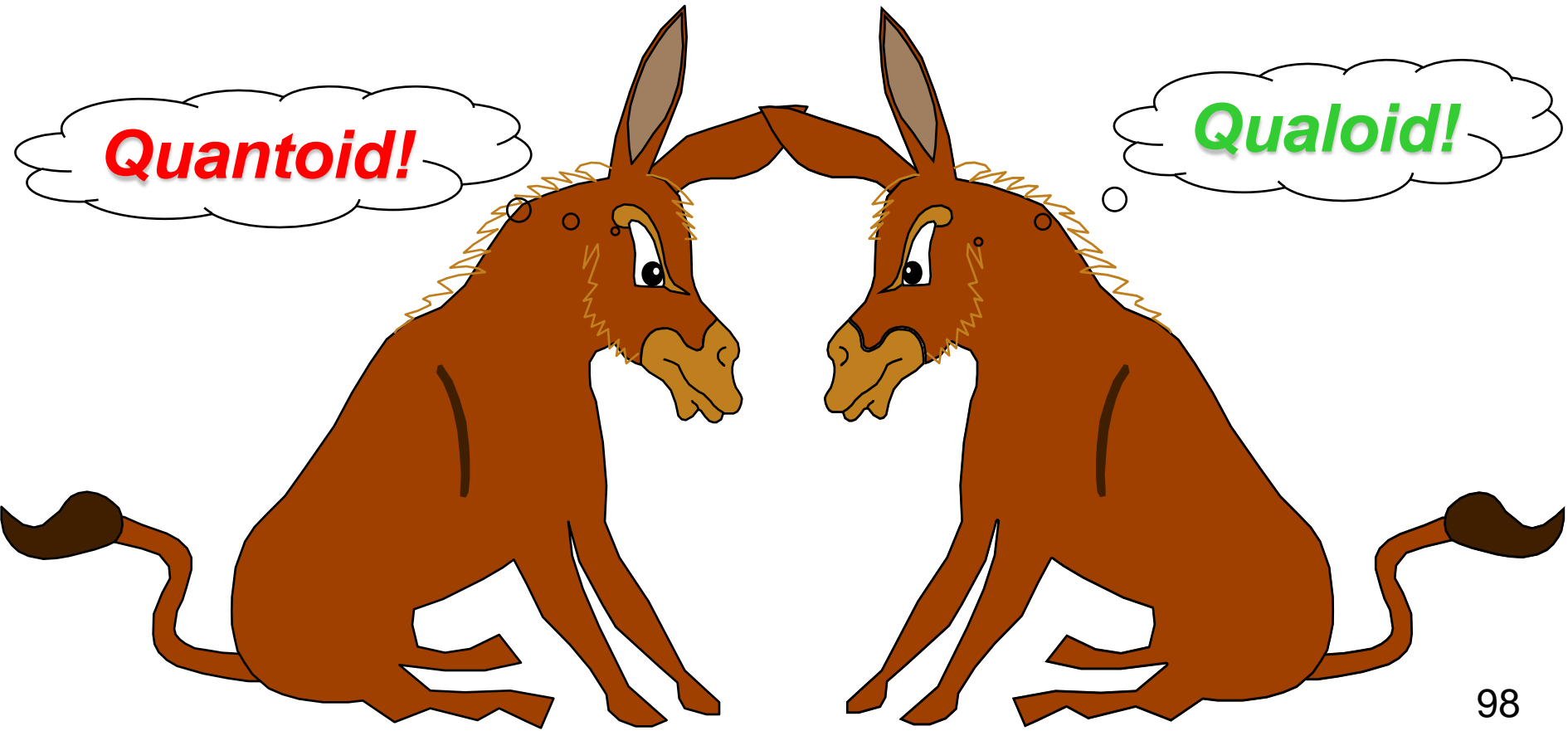
***Designing Evaluations under Budget,
Time, Data and Political Constraints***

**Mixed-method
evaluations**

It should NOT be a fight between pure

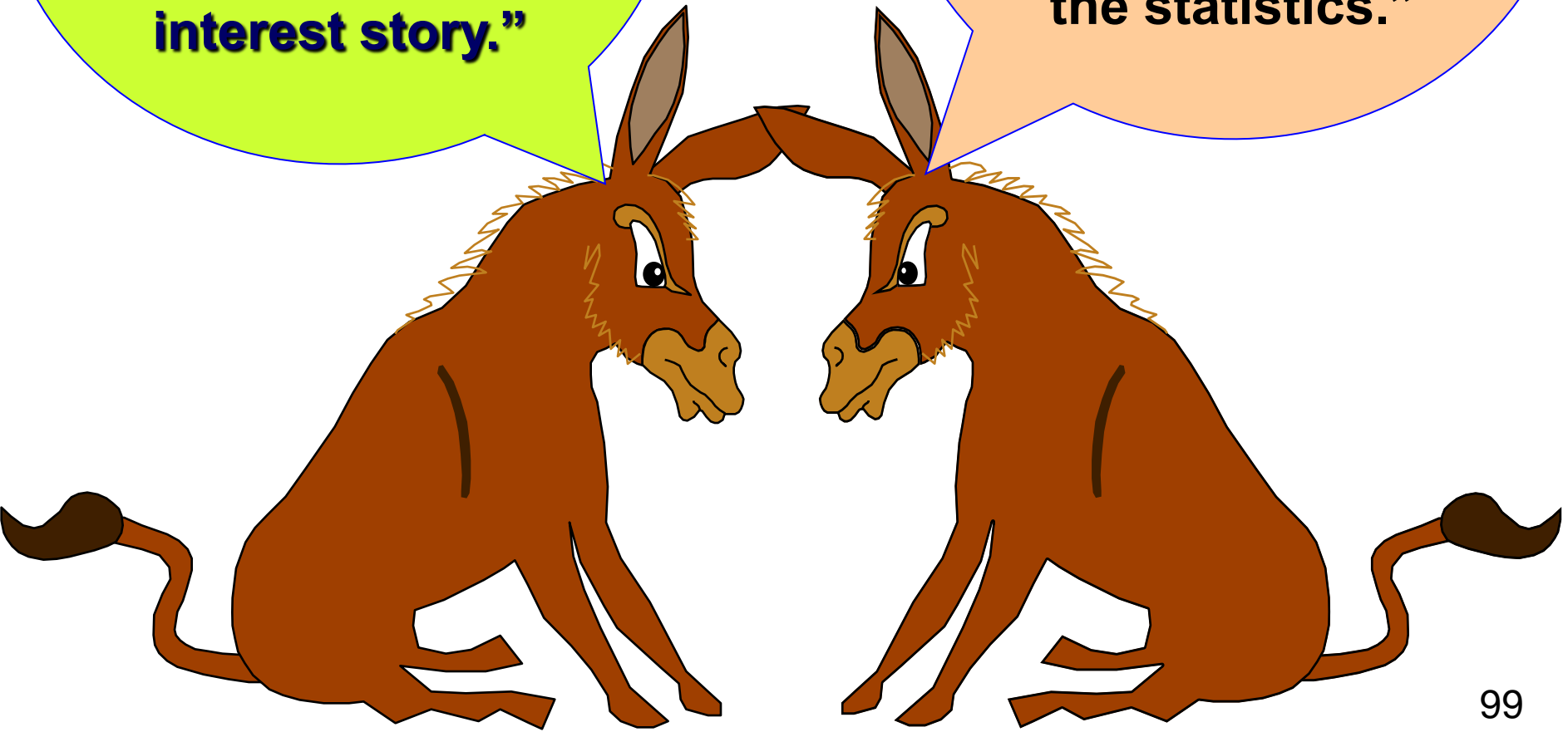
QUALITATIVE
(verbiage alone) OR

QUANTITATIVE
(numbers alone)

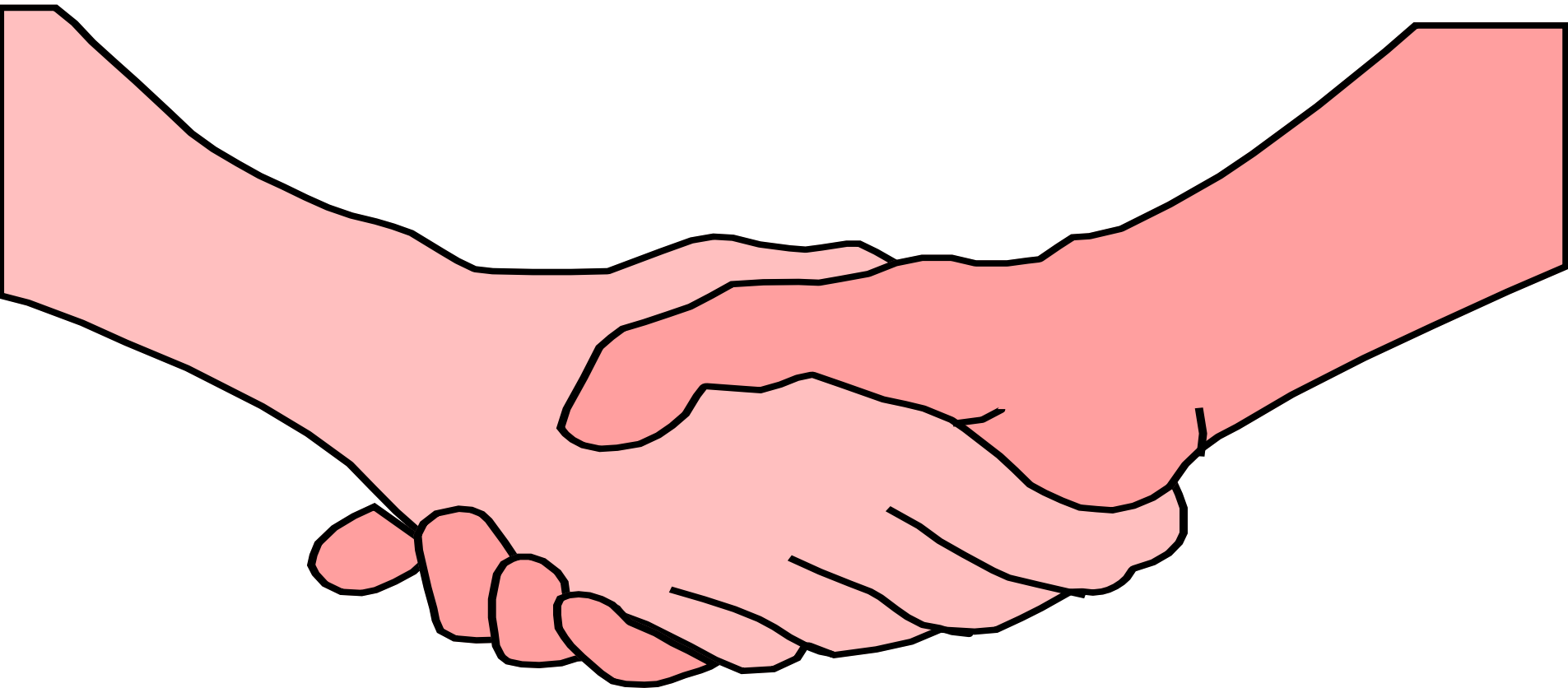


**“Your numbers
look impressive,
but let me tell
you the human
interest story.”**

**“Your human
interest story
sounds nice, but
let me show you
the statistics.”**



What's needed is the right combination of
BOTH ***QUALITATIVE methods***
AND ***QUANTITATIVE methods***



Participatory approaches should be used as much as possible



but even they should be used with appropriate rigor: how many (and which) people's perspectives contributed to the story?

Mixed method evaluation designs

How quantitative and qualitative methods complement each other

- A. Broaden the conceptual framework combining theories from different disciplines
- Exploratory **QUAL** studies can help define framework
 - **QUANT** surveys can measure extent of phenomena
- B. Combine generalizability with depth and context
- Random subject selection ensures representativity and generalizability
 - Case studies, focus groups, etc., can help understand the characteristics of the different groups selected in the sample

Mixed method evaluation designs

How quantitative and qualitative methods complement each other (cont.)

C. Permit access to difficult to reach groups

- PRA, focus groups, case studies, etc., can be effective ways to reach women, ethnic minorities and other vulnerable groups
- Direct observation can provide information on groups difficult to interview, e.g. informal sector and illegal economic activities

D. Enable Process analysis

- Observation, focus groups and informal conversations are more effective for understanding group processes or interaction between people and public agencies, and studying the organization

Mixed method evaluation designs

How quantitative and qualitative methods complement each other (cont.)

- E. Analysis and control for underlying structural factors [QUANT]
- Sampling and statistical analysis can avoid misleading conclusions
 - Propensity scores and multivariate analysis can statistically control for differences between project and control groups

Example:

- [QUAL] Meetings with women may suggest gender biases in local firms' hiring practices; however,
- [QUANT] using statistical analysis to control for years of education or experience may show there are no differences in hiring policies for workers with comparable qualifications

Example:

- [QUAL] Participants who volunteer to attend a focus group may be strongly in favor or opposed to a certain project, but
- [QUANT] a rapid sample survey may show that most community residents have different views

Mixed method evaluation designs

How quantitative and qualitative methods complement each other (cont.)

F. Triangulation and consistency checks:

- Direct observation may identify inconsistencies in interview responses
- Examples:
 - A family may say they are poor but observation shows they have new furniture, good clothes, etc.
 - A woman may say she has no source of income, but an early morning visit may show she operates an illegal beer brewing business

G. Broadening the interpretation of findings:

- Combining personal experience with “social facts”
- Statistical analysis frequently includes unexpected or interesting findings which cannot be explained through the statistics. Rapid follow-up visits may help explain the findings.

Mixed method evaluation designs

How quantitative and qualitative methods complement each other (cont.)

H. Interpreting findings

Example:

- A [QUANT] survey of community water management in Indonesia found that with only one exception all village water supply was managed by women.
- Follow-up [QUAL] visits found that in the one exceptional village women managed a very profitable dairy farming business – so men were willing to manage water to allow women time to produce and sell dairy produce.

Source: Brown (2000)

Using **Qualitative methods** to improve the Evaluation design and results

- Use recall to reconstruct the pre-test situation
- Interview key informants to identify other changes in the community or in gender relations
- Conduct interviews or focus groups with women and men to
 - assess the effect of loans on gender relations within the household, such as changes in control of resources and decision-making
 - identify other important results or unintended consequences:
 - increase in women's work load,
 - increase in incidence of gender-based or domestic violence



Questions?

Main workshop messages

- 1. Evaluators must be prepared for RealWorld evaluation challenges.**
- 2. There is considerable experience to draw on.**
- 3. A toolkit of rapid and economical “RealWorld” evaluation techniques is available (see www.RealWorldEvaluation.org).**
- 4. Never use time and budget constraints as an excuse for sloppy evaluation methodology.**
- 5. A “threats to validity” checklist helps keep you honest by identifying potential weaknesses in your evaluation design and analysis.**



Thank you!