Designing quality impact evaluations under budget, time and data constraints

BBL Co-sponsored by OED
and the Poverty Analysis, M&E Thematic Group

July 28, 2005

Michael Bamberger
Independent Consultant [mikejbamb@aol.com]

Assessing the impacts of livelihood programs

for the Bangladesh CHARS

I. Purpose of the presentation

To contribute to the Poverty Analysis, M&E Thematic Group and OED's efforts to improve the quality of impact evaluations by discussing situations in which researchers are asked to conduct impact evaluations

- under real-world constraints that limit the options for using rigorous evaluation designs
- but where there is a demand for acceptable estimates of project impacts.

- The presentation addresses the common situations under which researchers must decide whether and how acceptable impact evaluations can be designed, implemented, and used under real-world budget, time, data [and political] constraints.

Key questions

- How to define acceptable impact evaluation design standards?
  - What are the methodological cut-offs when impact evaluations are no longer possible?
- Options for addressing budget, time and data constraints
- Can mixed-method approaches strengthen randomized and “strong” quasi-experimental designs?

II. Conducting quality impact evaluation under budget, time, data and other real-world constraints

What is meant by quality impact evaluation?

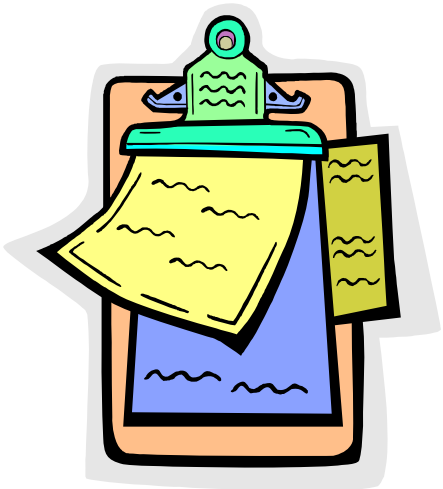
- Adequacy of response to client's information needs
- Methodological rigor
  - Assessed in terms of threats to conclusion validity, statistical power analysis, peer review etc
- Intervention outcome is assessed against an explicit counterfactual

Approaches to methodological rigor (poverty impact evaluation web site)

- Randomization or experimental designs
- “Strong” quasi-experimental designs
- Propensity score matching
- Pipeline comparison
- Simulated counterfactual
- Difference in means
- Difference in difference (double difference)
- Instrumental variables

Source: http://www1.worldbank.org/prem/poverty/ie/evaluationdb.htm

The real-world Evaluation approach [*a.k.a Shoestring Evaluation*]



An integrated approach to ensure acceptable standards of methodological rigor under real-world budget, time, data and political constraints.

RealWorld evaluation scenarios

Scenario 1. The evaluator is called in at the start of the project but for political, technical or budget reasons:

- It is difficult to collect data on control groups (if there is no control group we may no longer be talking about an impact evaluation unless we can develop a different counterfactual)
- Management may be reluctant to collect all of the required baseline data on future project participants

real-World evaluation scenarios

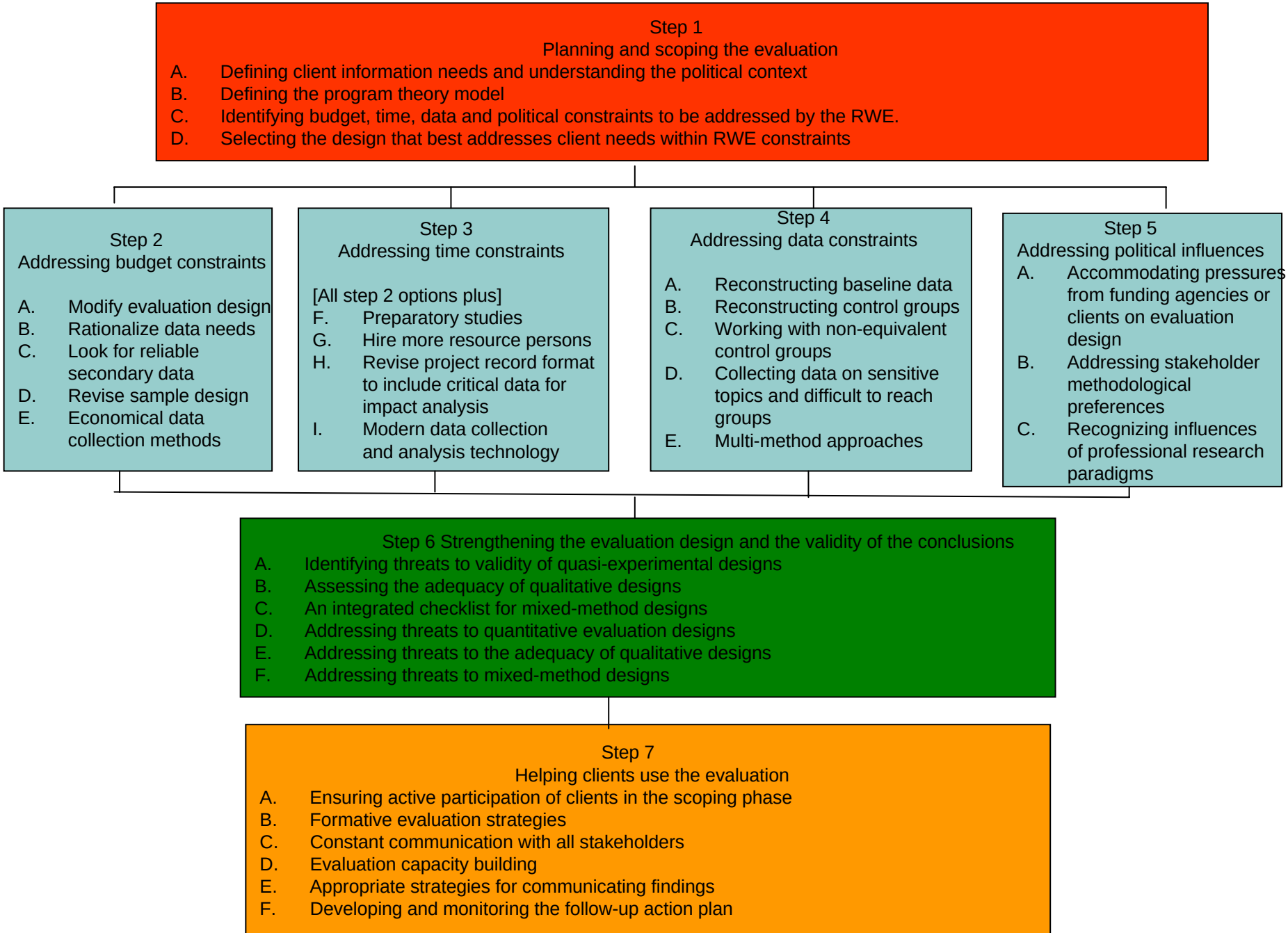
Scenario 2: The evaluator is not called in until the project is well advanced:

- No baseline or control group data has been collected
- Time pressures
- Budget constraints
- Political constraints on the evaluation methodology

Design approach varies depending on constraints

- **An adequate budget but lack of data**
- **A limited budget but plenty of time**
- **An adequate budget but limited time**
- **Political constraints**

The Real World Evaluation [RWE] Approach



Note:

- Not all steps of the approach are discussed in this workshop due to time constraints

Step 2

Addressing budget constraints

5 options

A. Simplify the evaluation design
B. Rationalize information needs
C. Using secondary data
D. Reducing sample size
E. Economical data collection methods

2A. Simplifying the evaluation design

The strongest general purpose Quasi-Experimental Design			
	T_1	T_2 Project intervention	T_3
Project group	P_1	X_1	P_2
Non-equivalent control group	C_1		C_2

Simplifying the evaluation design

- There are 2 methodologically strong QED [Models 1 and 2] and 5 less robust models that compensate for time, budget or data constraints by eliminating one or more of the 4 observation points.
- Each successive model is subject to more potential threats to the validity of conclusions.

Model 4: No pre-test control group

	Before project	Project implementation	End of project
	T_1	T_2	T_2
Project Group	P_1	X	P_2
Non-equivalent Control Group			C_1

Example: The effects of resettlement in the Maharashtra Irrigation Project, India

Model 5: No baseline data

	Before project	Project implementation	End of project
	T_1	T_2	T_3
Project Group		X	P_1
Non-equivalent Control Group			C_1

Example: Assessing the impacts of micro-credit on women and families in Bangladesh

Model 6: No control group

(is this still an impact evaluation?)

	Before project	Project implementation	End of project
	T_1	T_2	T_3
Project Group	P_1	X	P_2

Non-equivalent Control Group

Example: Impact of the Khao Laem Hydroelectric project in Thailand

2B. Rationalize data needs

Use Step 1 information on client information needs to eliminate any questions not directly related to client needs.

2C. Look for reliable secondary sources and assess their relevance and reliability

- Project planning studies
- Project administrative records
- Surveys by ministries and other government agencies
- Donor agency studies
- Universities
- Mass media

2D. Factors affecting sample size

1. **Purpose of the evaluation**

   *** Exploratory evaluations (does the program intervention "work") require smaller samples than quantitative hypothesis testing**

2. **Effect size**

   *** the smaller the expected effect the larger the required sample**

3. **Required level of precision**

   *** significant testing at 10% level (0.1) requires smaller sample than 5% level**

4. **Required power of the test**

*** with conventional Power = 0.8 (80% probability of detecting a significant difference if it really exists) the sample is much smaller than if a higher level of certainty (say Power = 0.9) is required.**

Ways to reduce sample size

- Increase effect size
- Accept lower Power
- Accept lower level of statistical precision
- Reduce the levels of disaggregation of the analysis
- Stratification or cluster sampling

2E. Reducing costs of data collection and analysis

- Self-administered questionnaires (be aware of potential bias)
- Reduce length and complexity of instrument
- Piggy-back on other survey (be aware of potential bias)
- Direct observation

Reducing costs of data collection and analysis

- Piggy-back on other surveys
- 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 foreign consultants
  - Commission preparatory studies
  - Video conferences
- Hire more consultants/researchers
- Incorporate impact indicators in project administrative documents
- Technology for data inputting/coding

Step 4: Addressing data constraints

- Lack of baseline data on project population
- Lack of control groups
- Statistical problems with non-equivalent control groups
- Collecting data on sensitive topics or from inaccessible groups

Reconstructing baseline data

- Using secondary data
- Recall
- PRA techniques
- Key informants
- Triangulation

Step 6: Threats to validity

4 sets of threats to the validity of conclusions from randomized and quasi-experimental designs

- Statistical conclusion validity
- Internal validity
- Construct validity
- External validity

5 sets of indicators for the validity and adequacy of mixed method designs

A. Confirmability and objectivity
B. Reliability and dependability
C. Internal validity, credibility and authenticity
D. External validity, transferability and fittingness
E. Utilization, application, action orientation.

Example: Selection bias (threat 2.2)

Project and control groups have different characteristics.

Possible solutions:

- Statistically control for differences
- Use secondary data to compare the two groups and identify differences
- Use key informants to explain differences
- Use direct observation to assess differences

Conclusion:
A. Ways to strengthen real-world evaluation designs

- Statistical matching of samples
- Process analysis to open-up the “black box” (project implementation process).
- Mixed-method designs
- Triangulation
- Using threats to validity checklist to identify and address design weaknesses

Conclusion
B. Cut-off points for acceptable impact evaluation

- Depends on purpose of the evaluation and the required level of precision
- Rigorous impact evaluation not always required
  - Important to define clearly the purpose of each study – is it a rigorous impact evaluation?
- Can threats to validity checklist be used to define cut-off points?

Conclusion
C: Strengthening randomized designs

Weaknesses of randomized designs

1. "Black-box" approach
   - Assumes project administered as planned and in uniform way
   - Assumes project treatment delivered at one point in time
   - Difficult to understand the process of change
2. Need to understand context within which project/programs implemented

Weaknesses of randomized designs

3. Standard measurements/indicators cannot be adapted to changes in project/context
4. Focus on quantitative but not qualitative dimensions

Addressing weaknesses in randomized designs

1. Process analysis:
   - Beneficiary assessment
   - Participant observation
   - Panel studies
   - Using monitoring data
   - Triangulation
2. Contextual analysis
   - Analysis of political, economic, administrative, environmental and socio-cultural variables through observation, PRA, key informants
3. Mixed-method approaches
   - Combining quantitative and qualitative approaches