Universidad de Los Andes

September 18, 2006

Expanding the Range of Tools for Increased Utilization of M&E in the Colombian Government

One Day Seminar

Addressing budget, time and data constraints in the design of impact evaluations

Michael Bamberger
[mikejbamb@aol.com]

Session outline

1. The trade-offs between in-depth and rapid evaluations.
2. The uses and limitations of randomized evaluation designs
3. Evaluation scenarios where rapid and economical methods are used
4. Design options for reducing costs and time
5. Data collection options for reducing costs and time

1: The trade-offs between in-depth and rapid evaluations

Trade-offs between in-depth and rapid evaluations

- In-depth pre-test/post-test control group designs can have a high degree of methodological rigor BUT
  - Expensive
  - Time-consuming
  - Sometimes lose credibility by not producing results for several years
- Rapid and economical evaluations can focus on priority issues and produce results when needed by planners and policymakers BUT
  - Methodologically less rigorous
  - Potential bias and increased risk of wrong conclusions
  - Lose credibility among rigorous quantitative audiences

Current debate in Colombia on the right mix of in-depth and rapid evaluations

- SINERGIA is producing high-quality impact evaluations but there is a concern that;
  - Expensive
  - Results sometimes too late to be useful (Empleo en Accion)
  - Delays affect credibility with decision-makers
- Methodologies for rapid evaluations (Evaluaciones Ejecutivas) lasting 3 months and costing US $10,000 – 15,000 are being developed to complement the in-depth impact evaluations.

Experience of the Ministerio de Hacienda in Chile on the mix of evaluations

- Each year Hacienda commissions an average of:
  - 14 rapid (3 month) evaluations
  - 4 in-depth impact evaluations.

- For a further discussion of these approaches and an assessment of their methodological limitations see:
  - Independent Evaluation Group. 2006. Conducting quality impact evaluations under budget, time and data constraints. World Bank

3. The uses and limitations of randomized designs

The uses and limitations of randomized designs

- The validity of findings and conclusions of an impact evaluation depend on the logical robustness of the counterfactual
- How well can alternative explanations of observed changes in the project population be eliminated?
- When they can be used randomized designs can produce the strongest counterfactual [see Handout 1 for examples of randomized designs].

Randomized designs are probably used in less than 5% of evaluations (see Handout 2) for the following reasons:

- Program design does not permit randomized selection of project beneficiaries:
  - Self-selection
  - Selected according to administrative criteria
- Methodological constraints
- Budget constraints
- Political constraints
- The evaluation does not begin until late in the project cycle

Limitations of randomized designs when used in isolation

- “Black box” approach
  - cannot distinguish between design failure and implementation failure
- Inflexible
  - Cannot adapt to changing project environments
- Normally relies on a few quantitative outputs indicators
- Ignores the unique political, economic, social and environmental factors in each project location.

- Most of the limitations can be addressed by using mixed-method approaches that combine randomized designs with qualitative methods such as:
  - Process analysis
  - Contextual analysis

3. Evaluation scenarios where rapid and economical methods are used

Evaluation scenarios when rapid and economical methods are often required

A. Assessing initial findings of pilot projects
B. Pre-intervention studies when conventional baseline studies are not possible
C. Process analysis to monitor the process of project implementation
D. Rapid feedback on problems identified by the monitoring system
E. To compliment in-depth evaluations during project implementation or at completion

[continues next page]

F. Post-intervention evaluations when baseline data has not been collected AND/OR

G. Post-interventional studies when working under budget and time constraints

H. Follow-up studies to fill in gaps in data analysis

I. Sustainability assessment when projects have been operating for several years

4. Evaluation design options to

reduce cost and time

Option 1 Using secondary data to reduce need to collect new data

- When available good quality secondary data can replace the need to collect pre or post-treatment data on the project and/or control groups.
- Permits the cost-effective use of more rigorous analytical methods (e.g. Propensity Score Matching).
- For most evaluations high quality secondary data is not available and cost-effect effective methods must be developed to collect pre and post-test data.

Table 1 Possible primary and secondary data sources for defining counterfactuals

Methodological validity	Primary data collection	Secondary data*
Strongest	Randomized design	
Strong	Strong quasi-experimental designs ● (Designs 1-2) with well matched pre and post-test control group	● Large sample surveys permitting the use of Instrumental Variables or Propensity Score Matching
Acceptable	● Evaluation designs 3-5. ● Baseline data or control group missing or weak	● National data sets providing control data for the total population but which do not identify the project group.
Weak or very weak	Evaluation designs 6 and 7. ● No control group and sometimes no baseline	● National data sets.

* Note: The adequacy of the secondary data (population coverage, quality, comparability etc) must always be assessed

Option 2: Simplify evaluation design

- Cost and time of data collection can be significantly reduced by eliminating one or more of the data collection points [see next slide]:
  - Baseline control group
  - Baseline project group
  - Post-test control group
- But this weakens the counterfactual and increases threats to the validity of the findings

Seven impact evaluation design options

[See Handout 4]

	T_1 Pre-project	Project begins	T_2 Implementation	T_3 End of project	T_4 Post-project
Design No.	**The two strongest evaluation designs**				
1	P_1 C_1	X	P_2 C_2	P_3 C_3	P_4 C_4
2	P_1 C_1	X		P_2 C_2	
	Three less robust but frequently adequate designs				
3		X	P_1 C_1	P_2 C_2	
4	P_1	X		P_2 C_1	
5		X		P_1 C_1	
	Two weak (non-experimental) but widely used designs				
6	P_1	X		P_2	
7		X		P_1	

Option 3: Modify sample selection strategy

- Stratified sampling can reduce total sample size if the within-strata variance is significantly less than between-strata variance.
- Cluster sampling by reducing distance between subjects can reduce average interview cost (but increases sampling error)

Option 4: Reduce sample size
[See Handout 5]

- Accept larger confidence interval for estimates (of mean etc)
- Accept lower confidence level (0.1 instead of 0.05 etc)
- Reduce level of disaggregation (see Option 5)
- Increase project effect size

Option 5: Simplify the outcome indicators

- Eliminate indicators requiring large samples, for example:
  - infant mortality rates, nutritional input, detailed expenditure surveys, anthropometric measures]
- Replace with less precise indicators, for example:
  - Morbidity, mothers recall of what children eat, recall of main expenditure categories
- This can significantly reduce required sample size but increases reporting error and reduces precision of outcome estimates

Option 6: Simplify the analysis plan

- Reduce the levels of disaggregation of the analysis. For example:
  - Estimate impacts for total project population and not by region, project site, types of household etc.
- Simplify the definition of the target and control groups
- This will often permit a significant reduction in the sample size.

5. Data collection options for

reducing time and cost

Option 1: Reduce the amount of information to be collected

Consult with the following sources and eliminate information not required to answer priority questions:

- clients and stakeholders
- Key informants and experts
- program theory model
- Analysis plan

Option 2: Effective utilization of secondary data

- Always check availability of secondary data before collecting more primary data
- Multiple secondary data permits triangulation and increases reliability/validity

[continues next page]

- Assess the appropriateness of secondary sources before using
  - Timing
  - Population coverage
  - Who interviewed?
  - Coverage of key indicators
  - Quality/reliability of data collection

Option 3: Collect data at higher level of aggregration

- Collect:
  - community rather than household level data
  - Collect school rather than student level data
- Reduces cost but also changes unit of analysis from individual to community/group so the sample size becomes much smaller

Option 4: Replace sample surveys with observation

- Observation checklists can cover
  - Travel and transport patterns (walking, bus, animal traction etc)
  - Time-use patterns (collecting water and fuel)
  - Sexual division of labor (particularly in agriculture)
  - Economic status of the community (quality of house construction, social infrastructure, no. of vehicles, quality of clothing etc)

Option 5: Replace baseline data with recall

- Recall can provide estimates of pre-intervention conditions when baseline surveys were not conducted:
  - Access to school and clinics
  - Access to transport and costs
  - Water utilization
  - Social organization and conflict
  - Fertility behavior
  - Income and expenditures

[continued next slide]

- Very few studies or guidelines to assess recall reliability, validity and bias:
  - Expenditure and fertility two areas with research literature
- Expenditure literature shows results sensitive to how questions asked
  - Recall period
  - How data recorded
  - Number of expenditure categories
- Important to use triangulation to strengthen reliability
- For agencies such as DNP that conduct large scale longitudinal studies it would be easy to test the validity of recall

Option 6: PRA techniques

- Participatory assessment methods collect data from groups or communities and can be used to:
  - Assess economic status and stratification
  - Access and quality of services such as water and sanitation
  - Changes over time (seasonal or longer periods, trend analysis)
  - Causal analysis

Examples of PRA techniques for reconstructing historical data

- Time line
- Trend analysis
- Historical transect
- Seasonal diagram
- Daily activity schedule
- Participatory genealogy
- Dream map
- Critical incidents

Source: S, Kumar 2002 Methods for community participation. ITDG publications.

Option 7: Reducing costs of survey data collection [see Handout 6]

- Use less expensive data collectors:
  - Medical students, university students, high school students, community groups
  - Problems of quality, reliability and bias
  - Hidden costs as more training and supervision required
- Providing respondents with mobile phones to reduce interviewer travel time [also useful when there are security problems]
- Hand-held computers to reduce costs of data input and analysis