

RealWorld Evaluation

Designing Evaluations under Budget, Time, Data and Political Constraints

Facilitated by **Jim Rugh**

Note: The PowerPoint presentation, condensed summary chapter of the book and other resources are available at: www.RealWorldEvaluation.org

Workshop Agenda

1. Introduction [10 minutes]
2. Brief summary of the RealWorld Evaluation (RWE) approach [30 minutes]
[slides 1-21]
3. **Small group self-introductions and sharing of RWE issues you have faced in your own practice. [20 minutes]**
4. RWE Step 1: Scoping the evaluation, including evaluation design, counterfactuals, etc. [30 minutes] [slides 22-77]

[short break]

5. RWE Steps 2, 3 and 4: Addressing budget, time and data constraints [60 minutes] [slides 78-111]
6. **Small groups prepare their case studies (60 minutes)**

[lunch break 60 minutes]

7. Quantitative, qualitative and mixed methods [20 minutes] [slides 112-131]
8. **Small groups meet in pairs ('consultants' with 'clients' to negotiate their ToRs (in case studies) (60 minutes)**
9. **Optional session (time permitting): Small groups discuss how what they have learned will help them address constraints identified in introductory small-group discussion. [25 minutes]**
10. Wrap-up discussion, evaluation of the workshop [15 minutes]

RealWorld Evaluation learning exercise

Discussion Group A (Evaluators)

Designing an impact evaluation for a housing project under **budget** constraints



A four-year pilot housing project is being launched this year for low-income families in 5 neighborhoods of a large city. The project and its evaluation are being funded by a major foundation and will be implemented by the City Housing Department. The project objective is to improve economic and social conditions of poor families through good quality and affordable housing. Any family with an income below 150% of the poverty line is eligible to apply. Successful applicants will be selected by the Housing Department. It is expected that at least 20,000 families will apply, out of which housing units will be provided to 5,000 families during the four-year pilot project

If this pilot phase is successful the possibility of a second and larger project would be considered by the foundation, and the findings of the evaluation would be an important factor in deciding whether to approve the second project. A consultancy evaluation team has submitted a preliminary proposal for an evaluation design to assess the impacts of the project. They proposed using a pretest - posttest comparison of the project population with a sample of households in comparable areas of the city not served by this project. A household sample survey would be administered in early in year one and again in year 3 to a sample of 1,500 project households covering all 5 project locations, and 1,500 comparison group households. The proposed evaluation design is given below.

The proposed pre-test post-test quasi-experimental evaluation model [Design 2]			
	T ₁ (baseline)	T ₂ Project intervention	T ₃ (year 3)
Project group covering all 5 project locations [n = 1,500]	P ₁	X	P ₂
Comparison group representing the types of housing where project families previously lived [n = 1,500]	C ₁		C ₂

The foundation acknowledged the proposed evaluation design was technically sound but indicated that the proposed cost exceeded their budget. The Housing Department, which has little experience with formal evaluations, felt the evaluation was too complicated and expensive.

For the evaluation of their previous projects the Department contracted an architect who had visited the project sites, interviewed project staff, local officials and a few project beneficiaries. In the opinion of the Housing Director this system had worked well and they did not see any reason to change it. The Director is also concerned that interviewing a comparison group might “create problems”. The consultants were asked to resubmit the proposal with the budget reduced by at least 25 per cent and if possible by 50 per cent.

The Exercise

Imagine that you are the members of the **consultancy evaluation team** being considered to be given the contract to develop this evaluation plan and then subsequently conduct it.

Question 1. What will you do to address the constraints presented, i.e. reduce the cost of the proposed evaluation plan by 25% to 50%?

Question 2. More specifically, which one or combination of the RealWorld Evaluation options for addressing **budget** constraints would you propose for this evaluation? (You can select more than one.)¹

- A. Simplify the evaluation design (describe how)
- B. Clarify client information needs
- C. Look for reliable secondary data
- D. Reduce the sample size
- E. Use more economical methods of data collection
- F. Other (describe)

Question 3: What are the trade-offs in the options you are considering? How would they be addressed in your revised evaluation plan and/or in your negotiations with the clients (Housing Department and the Foundation)?²

Question 4: Identify some of the political factors that might influence how the evaluation is designed and implemented. What are likely to be the major concerns of the foundation? What are the concerns of the Housing Department? Why is the Director worried about interviewing a comparison group? How would you try to convince them of the limitations of their traditional way of conducting evaluations? How would you address the Housing Department’s concerns about the use of a comparison group? Do you have any concerns that there are sensitive political issues and that the clients may pressure you to avoid these or to handle them in a certain way? How would you address these pressures?

¹ *Suggestion: Refer to the “Choosing the Best Design from the Available Options” section on pp. 12-20 in the Condensed Overview of RealWorld Evaluation, and also tables 1 and 2.*

² *Suggestion: Refer to Table 5: Condensed Checklist For Assessing The Adequacy And Validity Of All Evaluation Designs (pp.49 ff). See especially sections B, C and E.*

RealWorld Evaluation learning exercise

Discussion Group B (Clients)

Designing an impact evaluation for a housing project under **budget** constraints



A four-year pilot housing project is being launched this year for low-income families in 5 neighborhoods of a large city. The project and its evaluation are being funded by a major foundation and will be implemented by the City Housing Department. The project objective is to improve economic and social conditions of poor families through good quality and affordable housing. Any family with an income below 150% of the poverty line is eligible to apply. Successful applicants will be selected by the Housing Department. It is expected that at least 20,000 families will apply, out of which housing units will be provided to 5,000 families during the four-year pilot project

If this pilot phase is successful the possibility of a second and larger project would be considered by the foundation, and the findings of the evaluation would be an important factor in deciding whether to approve the second project. A consultancy evaluation team has submitted a preliminary proposal for an evaluation design to assess the impacts of the project. They proposed using a pretest - posttest comparison of the project population with a sample of households in comparable areas of the city not served by this project. A household sample survey would be administered in early in year one and again in year 3 to a sample of 1,500 project households covering all 5 project locations, and 1,500 comparison group households. The proposed evaluation design is given below.

The proposed pre-test post-test quasi-experimental evaluation model [Design 2]			
	T ₁ (baseline)	T ₂ Project intervention	T ₃ (year 3)
Project group covering all 5 project locations [n = 1,500]	P ₁	X	P ₂
Comparison group representing the types of housing where project families previously lived [n = 1,500]	C ₁		C ₂

The foundation acknowledged the proposed evaluation design was technically sound but indicated that the proposed cost exceeded their budget. The Housing Department, which has little experience with formal evaluations, felt the evaluation was too complicated and expensive. For the evaluation of their

previous projects the Department contracted an architect who had visited the project sites, interviewed project staff, local officials and a few project beneficiaries. In the opinion of the Housing Director this system had worked well and they did not see any reason to change it. The Director is also concerned that interviewing a comparison group might “create problems”. The consultants were asked to resubmit the proposal with the budget reduced by at least 25 per cent and if possible by 50 per cent.

The Exercise

Imagine that some members of your group represent the **Housing Department** and others represent **the Foundation** that is funding the project. You are having a meeting together to consider the initial proposal submitted by the consultancy group. In light of what you have learned about the RealWorld Evaluation approaches to evaluation, you are preparing to negotiate the Terms of Reference (in broad terms, not in detail) with the consultants. Note: the two agencies will have different approaches to most of the following questions and you must try to understand their different perspectives.

Question 1. What will you propose be done to address the constraints, i.e. reduce the cost of the proposed evaluation plan by 25% to 50%?³

Question 2 What are the trade-offs in the options that are being considered? How would they be addressed in your revised evaluation plan and/or in your negotiations with the evaluation consultants?⁴

Question 3. Summarize the major challenges to conducting what each agency would consider an acceptable impact evaluation when there are **budget** constraints of these kinds. What are the differences in their expectations for the evaluation? What are some suggestions for how the constraints might be addressed and how can the differences in the expectations of the two agencies be reconciled?

Question 4. What are some of the main points that each agency would like to emphasize in their negotiations with the consultants? What are some of the political issues that might influence how the evaluation is designed and implemented? Are there any sensitive issues that the evaluation should not discuss, or that must be addressed in a certain way? What advice would each agency give to the consultants on these issues?

³ *Suggestion: Refer to the “Choosing the Best Design from the Available Options” section on pp. 12-20 in the Condensed Overview of RealWorld Evaluation, and also tables 1 and 2.*

⁴ *Suggestion: Refer to Table 5: Condensed Checklist For Assessing The Adequacy And Validity Of All Evaluation Designs (pp.49 ff). See especially sections B, C and E.*

Alternatives to the Conventional Counterfactual⁵

Facilitators: Michael Bamberger [jmichaelbamberger@gmail.com],

Fred Carden [fcarden@idrc.ca] and Jim Rugh [jimrugh@mindspring.com]

(There were about 50 participants in this session)

The purpose of this note is to provide a summary of the main points raised during the discussion. It is hoped this will stimulate a continued dialogue and sharing of experiences. The Think Tank had three objectives:

- To ask the question: “How can project or program outcomes and impacts be assessed when it is not possible to construct a statistically matched control group?”
- To consider the real-world challenges of using comparison group designs and
- To share experiences on promising approaches for assessing outcomes without a statistical comparison group

Introduction:

The purpose of a counterfactual in evaluations is to address the question “What would have been the situation of the project population if the project had not taken place?” One method for answering that question is to conduct a Randomized Control Trial (RCT) or to use quasi-experimental designs with a pretest/posttest comparison of the project and a matched control group. However, the complexity of many community interventions, RealWorld budget, time and data constraints, and the fact that many evaluations are not commissioned until late in the project cycle, mean that conventional statistical evaluation designs are often not feasible and **alternatives to the conventional statistical counterfactual are an essential tool in the evaluator’s arsenal**. The increasing demand on evaluators to assess general budget support and complex, multi-donor, national-level programs where project level evaluation designs can rarely be applied further increases the need for alternative ways to assess the outcomes and impacts of these complex programs.

The following questions were addressed in the Think Tank:

- What alternatives are there to the conventional, statistical counterfactual?
- When and how can they be used?
- How adequately can they assess/eliminate alternative explanations of the changes observed in the target population?
- Which of the approaches can be used to assess the outcomes of complex, national level and multi-donor programs?

⁵ [Summary of Session 713 Think Tank during AEA 2010 conference in Orlando](#)

Summary of the points raised in the discussion

Table 1 summarizes a range of approaches that have been used or proposed to define the counterfactual and combines the initial list distributed in the workshop with the additional approaches proposed by workshop participants. Part A lists some of the conventional statistical approaches and Part B list alternative approaches that can be used when statistical designs with matched samples cannot be used. The approaches in Part B are divided into: B-1: theory based approaches, B-2: quantitatively oriented approaches and B-3: qualitatively oriented approaches. It should be noted that there is considerable overlap between the categories as many approaches combine two or more of the categories.

1. The challenges facing attempts to estimate project outcomes or impacts under RealWorld evaluation constraints

a. Many evaluations are conducted under budget and/or time constraints that often limit the numbers and types of respondents who can be identified and interviewed. The kinds of respondents that can most easily be identified and are easiest to interview are often project beneficiaries. Consequently many evaluations have a *systematic positive bias* as most of the information is collected from people who have benefited from the project. This is a common scenario where there is no plausible counterfactual representing the situation of people who were not affected by the project or who might even be worse off [see Box 1].

Error! Reference source not found.

b. Often evaluators have to find creative ways to identify sources of information on communities, groups or organizations that have similar characteristics to the project population except that they were not exposed to the project services or benefits. Examples of comparison groups include the following:

- i.* Schools close to, but outside the project area
- ii.* Health clinics serving those outside the target community
- iii.* Cooperative markets that serve farmers from project areas and from similar agricultural areas without access the project
- iv.* Villages without access to water or sanitation.

Error! Reference source not found.

c. Many participants in this Think Tank session expressed concerns about the dangers of using method-driven approaches to assessing outcomes and impacts. These approaches are considered to be supply driven where many evaluations are designed around the methodological requirements of particular designs rather than starting from an understanding of the problem being addressed and the unique characteristics of the program being studied. As one participant put it “A focus on methods leads us to look for the lost key under the streetlamp”. In other words we measure what is easy to measure rather than what needs to be measured. How to move the thinking on evaluation beyond the attempt to approximate the experimental design paradigm?

2. Examples presented by workshop participants of creative ways to identify comparison groups

The following are some of the approaches that participants had used [organized in terms of the categories in Table 1]. The classification is somewhat arbitrary and we indicate that many of the approaches could fall into several categories.

B-1: Theory-based approaches.

a. *Process tracing* looks at the little steps along the way and asks “Could there have been another way of doing each step, and what difference would this have made?” The evaluator can look to see if some of the alternatives can be identified. It will often be found that even when a project is intended to be implemented in a uniform way, there will often be considerable variations – sometimes through lack of supervision, sometimes because implementing agencies are encouraged to experiment and sometimes due to circumstances beyond the control of implementing agencies (such as lack of supplies or cuts in electrical power). Cultural differences among the target communities can also result in varieties of implementation. [Also B-3]

*** The practical lesson is that the evaluator must always be on the look-out for these variations as they offer the evaluator a set of ready-made comparison groups. However, these variations will only be found if the evaluator is in close and constant contact with the areas being studied and if the evaluation design has the flexibility to take advantage of variations as they appear.*

b. It is possible to learn from the methods of historical analysis to examine questions such as “What if a certain law had been approved earlier? What difference would it have made?” [See table 1]. For example, a hypothetical historical analysis was made to estimate how many lives of young children might have been saved if Megan’s Law had been implemented earlier. It was estimated that in 1 out of 25 cases studied a pedophile would have been detected before another offense was committed.

*** Similar approaches could be used for assessing the changes that had resulted from introducing a new project or service, by asking “What if the service had not been introduced?” and then trying to find similar groups or communities where the program was not operating.*

c. Another participant described the compilation of a “book of causes” based on observations, information interviews etc that lists possible causes of the observed changes in the lives of community households (including the effects of the project). The evaluator then returns to the field to assess the potential effects of the project by finding evidence to eliminate (or not eliminate) alternative explanations. [Also B-3]

d. An important, but often overlooked approach is to ask project families “What did the project do? What if any changes took place in the life of your family or the community? Would you attribute any of those changes to the effects of the project?” It is important not to assume that the project did have impact and to avoid asking the questions in such a way that the respondent feels obliged to agree that there were impacts.

e. Use a Venn diagram: Ask the community or group about the changes that have occurred, and then diagram those institutions or interventions that they feel had significant (large circles) or minor (small circles) influence on those changes, and how much (or little) they overlap with the community itself. The point is to facilitate this PRA activity objectively, without biasing the responses to the specific project.

f. A related technique is *time-change mapping*. In Ghana ILAC asked: “What have **we** done? Have we changed the perspective to that of the organization? What has been the evolution of this organization over a long period of time and who has contributed what to that?” The analysis was based on organization records - contracts, legal agreements - to map out change over time. See where their organization showed up. They saw a co-evolution of the objectives of the objectives of the different partners.

g. In this example the comparison is time. The structure of an organization is compared at different points in time and an analysis is made based on all available sources of information to determine what contributed to the observed changes.

h. Another different approach is to use *abduction: logic of discovery* dealing with the new scientific facts to work in the messy world. When we see something new this guides us to convert our conclusions into evidence.

i. Several participants referred to the potential utility of the forensics model but one person questioned how widely this approach can be used in program evaluation. “The forensic pathologist has a corpse; but in many development programs there is no body; all there is is an assertion; so the sequence of questions makes a big difference and the first step is to find out what happened because the documentary evidence is often inadequate or fallacious. It was pointed out that this is particularly true when evaluating complex systems or country-level programs. [Also B-3]

B-2: *Quantitatively oriented approaches*

a. An evaluation was conducted in Uganda to assess the impacts of a USAID HIV/AIDS program operating in 16 districts when only 3 weeks was authorized for data collection for the evaluation. It was decided to use the next closest district where the HIV/AIDS program was not operating as the comparison group. As often happens, it was found that there were no differences in outcome between the project and comparison groups, because a larger HIV/AIDS program was operating in many of the comparison areas.

*** This illustrates a common methodological challenge for evaluations: how to control for the effects of other programs operating in the same areas, or affecting the comparison groups?*

b. Rather than comparisons between areas with and without projects, many agencies are more interested to compare their project with alternative interventions. This approach was used by CARE Zambia to evaluate a school program [also B-3].

c. In another state-wide education evaluation in Chennai, India, the project covered all schools so it was not possible to identify a “without” group. But useful comparisons were made between different dosages of the treatments, and between outcomes for pilot projects and when

the program was rolled out to all schools in the state. One interesting determinant of outcomes was differences in motivation of participants to succeed.

d. It was pointed out that the designs for evaluating pilot projects will often be different from the designs used to evaluate scaled-up projects. The fact that pilot projects usually only affect a small proportion of the population means that possibilities exist for selecting a “without” comparison group that are will not be there for evaluating the state-wide scale-up.

*** Many participants were involved in projects or programs that covered all or most of the population so that the option of selecting comparison groups not affected by the project does not exist. Consequently a number of creative options must (and in practice can) be found. The following are some further examples.*

e. Participants mentioned similar approaches that compare the implementation of projects in different regions or where the outcomes of similar treatments were different. For example: evaluations of agricultural projects often compare differences in climate, how farms are organized, or other contextual differences.

f. Many evaluations take advantage of natural variations (sometimes called natural experiments). Sometimes these may be climatic variations; in other cases they may unexpected delays in launching the project in some areas (so there is a temporary “without” situation). In other cases the variations may be due to lack of supplies or trained staff so that some schools may only get the new textbooks but not the specially trained teachers, whereas other schools may receive specially trained teachers but not the new textbooks [also B-3].

g. The natural variations can be used to suggest patterns that can then be used to go beyond what can be seen to develop program theory models [also B-1 and B-3].

h. These natural variations can sometimes be used to implement a *pipeline evaluation design* whereby the 2nd cohort can serve as a comparison group to the 1st cohort, etc.

i. Another kind of time comparison is *cohort analysis*. In the BMGF project in Zambia several cohorts of people (each cohort entered the project in a different year) were interviewed at different points in time and the question was asked “How many people have additional sources of income?” Respondents were asked to compare their income before, during and after the project. A comparison was made between the experience of different cohorts and between participating cohorts and between cohort members and the fourth family to the north (not benefiting from the project but close enough to have similar characteristics). The evaluation design also built in controls (such as triangulation among estimates from different sources) for recall bias.

j. For this and similar kinds of analysis it is always important to interview key informants not directly involved, community leaders, government officials, etc. Also in each village try to interview at least a couple of families who had not participated in the project.

*** This seemingly simple and obvious step proves to be very important as evaluators working under budget and time pressures often end up mainly interviewing beneficiaries, and as we saw earlier the*

findings of the evaluation will often have a positive bias.

B-3 Qualitatively-oriented approaches

- a. Sometimes it is possible to identify and study projects operating with pre-selected different combinations of treatments, outcomes or external factors, but in many other cases these differences are identified in group discussions with, for example, farmers who are asked to describe different farming practices, ways that farms are organized etc. If farmers are asked both about their own experiences and their knowledge of other farmers (often including relatives or friends in different regions) it is often possible to identify quite a wide range and combination of factors.
- b. Other qualitatively oriented approaches referred to in earlier sections include:
 - a. Comparing projects with alternative interventions
 - b. Comparisons among projects offering different combinations and levels of treatments
 - c. Taking advantage of natural variations
 - d. Process tracing
 - e. Compiling a “book of causes”
 - f. Applying the forensics model
 - g. Time-change mapping

Table 1 Alternative strategies for defining a counterfactual when statistical matching of samples is, and is not possible.		
Approach	Description	Applications and issues
A. CONVENTIONAL APPROACHES FOR DEFINING A COUNTERFACTUAL THROUGH STATISTICAL MATCHING		
a. Randomized control trials	Random assignment of subjects to treatment and control groups with before-and-after comparison	Important to note that even though subjects are randomly assigned it is normally not possible to control for external factors during project implementation, so this is not a truly 'controlled' experimental design.
b. Quasi-experimental designs with statistical matching	Non-random assignment but with subjects matched using statistical techniques such as propensity score matching and concept mapping	
c. Quasi-experimental designs with judgmental matching	Subjects are matched judgmentally combining expert advice, available secondary data, and sometimes rapid exploratory studies.	
d. Regression discontinuity	A selection cut-off point is defined on an interval or ordinal scale (for example, income, expert rating of subjects on their likelihood of success or their need for the program). Subjects just above the eligibility cut-off are compared with those just below after the project treatment using regression analysis	Can provide unbiased estimates of treatment affects as long as the cut-off criterion is rigorously administered. In practice it has proved difficult to satisfy this requirement.

**B. ALTERNATIVE APPROACHES FOR CONSTRUCTING A COUNTERFACTUAL
WHEN STATISTICAL MATCHING OF SAMPLES IS NOT FEASIBLE**

B-1: Theory based approaches

a. Theory of change	A theory of change approach identifies what change is intended, how your program or project will contribute and how you will know you are making progress towards that change. A number of different retrospective approaches have been developing including <i>inter alia</i> Most Significant Change, Outcome Mapping.	Can be used at the project or program level where a program or project is deliberately seeking to create social change. There is also potential use at the systems level. May be comparative across projects or programs or comparative within a single project or program. Value is enhanced when the approach is treated as emergent, that is, incremental change is monitored and contributes to modifying the theory of change over time.
b. Program theory/ logic models	Program theory models describe the process through which a program is prospectively intended to produce changes, how changes will be measured, and contextual factors that might explain variations in outcomes in different locations. Results chains can also identify some of the potential negative effects/outcomes at different points in the project. At the end of the project theory models can be used retrospectively to assess the extent to which observed outcomes conformed to, or deviated from the theoretical model, and thus can be used to assess confidence in the validity of the model.	Can be used wherever a clearly defined program implementation strategy can be defined. Easier to apply at the project level but can also be used to evaluate complex, multi-component programs – although this will often require that programs are broken down into individual and more easily described and assessed components. The value for comparative analysis is greatly enhanced when one or more alternative theory models are defined to test rival hypotheses.
c. Realistic (sometimes called “realist”) evaluation	The approach focuses on the specific context in which a program is implemented and addresses the questions: “What works?”, “For whom?”, “When?” and “Why?” Whereas the conventional	Used where understanding of context is a critical component of assessing impact. This will often be the case when a program is operating in a number of different locations and there are differences in outcomes even when the program is supposedly implemented in the

	counterfactual focuses on the comparison group and often pays very little attention to the “factual”, realist evaluation seeks to understand in detail the mechanisms through which an intervention operates and how different sectors of the target population are affected.	same way in all locations. Also useful when decisions have to be made on the potential for replication in different settings or where the design and implementation strategies may be modified.
d. Asking participants “What was our project?”	Participants are also asked “What effects did it have and did it make any changes in the life of your family or the community?”	It is important not to assume that the project did have impact and to avoid asking the questions in such a way that the respondent feels obliged to agree that there were impacts.
e. Process tracing	Looks at the little steps along the way and asks “Could there have been another way of doing each step? And what difference would this have made?”	The evaluator can look to see if some of the alternatives can be identified. It will often be found that even when a project is intended to be implemented in a uniform way, there will often be considerable variations – sometimes through lack of supervision, sometimes because implementing agencies are encouraged to experiment and sometimes due to circumstances beyond the control of implementing agencies (such as lack of supplies or power cuts). Cultural differences among the target communities can also result in varieties of implementation.
f. Venn diagrams	The community is asked about changes that have occurred without orientation to a specific project	The process involves ascertaining the community’s own perspectives as to what institutions or factors contributed to those changes.
g. PRA time-related techniques [see also qualitative techniques]	Community or other groups work together to create diagrams or charts to describe the changes that have occurred in the community (for example, economic, environmental or organizational changes). Sometimes they are asked to link these to critical events (including the launch of a project) that might have influenced the changes.	In Ghana ILAC asked: “What have we done? Have we changed the perspective to that of the organization? What has been the evolution of this organization over a long period of time and who has contributed what to that?” The analysis was based on the organization’s records (including contracts, legal agreements) to map out change over time. See where their organization showed up. They saw a co-evolution of the objectives of the different partners.

h. PRA relational techniques [see also qualitative techniques]	Community or other groups work together to create diagrams or charts to describe relations between important events. This may be done through techniques such case-and-effect diagrams, impact diagrams, Venn diagrams or process maps.	The techniques can be used to assess the influence/impacts of projects or other external interventions.
i. Historical methods	Political, economic and social historians often try to reconstruct hypothetical scenarios to assess the impact of certain phenomena by asking the question “What would have been the consequences if a particular event had not taken place.	One example was a study that tried to assess the impacts of the construction of the US railroad system in the 19 th and early 20 th century. In another example, a hypothetical historical analysis was made to estimate how many lives of young children might have been saved if Megan’s Law had been implemented earlier. It was estimated that in 1 out of 25 cases studied a pedophile would have been detected before another offense was committed.
j. Forensic methods	Various writers, including Michael Scriven, have suggested that lessons can be learned from how forensic scientists determine the cause of death by searching for contextual clues and then working backwards through time after a corpse has been discovered.	While this approach is attractive there is a question about how widely it can actually be used in program evaluation. Many program evaluations do not even have a “body” from which the forensic approach could begin.
k. Criminal investigation methods	Police investigators piece together clues to identify the most likely explanation of a death by testing a hypothesis and then seeking to eliminate alternative explanations to determine the probable cause “beyond a reasonable doubt”.	
l. Drawing lessons from fields outside the social sciences	Other professions such as law, chemistry, theology, architecture and astronomy use very different approaches for assessing causality	We have not yet explored these fields but there may be some interesting lessons for evaluators.

m. Compilation of a “book of causes”	Based on observations, information interviews etc list possible causes of the observed changes in the lives of community households (including but not limited the effects directly attributable to the project).	The evaluator then returns to the field to assess the potential effects of the project by finding evidence to eliminate (or not eliminate) alternative explanations.
B-2 Quantitatively oriented approaches		
a. Pipeline design	When programs are implemented in phases over a period of time, the segments of the population only affected by the latter phases can be used as comparison groups for the earlier phases.	Widely used for evaluating large-scale infrastructure projects such as water or roads. The approach requires that: (a) the populations covered by each phase have similar characteristics; (b) the populations in later phases do not receive benefits until the scheduled start of their phase. Possible applications for evaluating sector or national programs include: (a) programs (for example, anti-corruption or decentralization) that will be implemented in different ministries over a period of time; (b) programs that will be implemented in different regions over a period of time; (c) programs that are intended to be implemented throughout the country but where there are implementation delays in some regions.
b. Natural variations (a variant of pipeline designs)	Taking advantage of climatic variations, unexpected delays in launching the project or variations due to lack of supplies or staff in some areas.	
c. Creative uses of secondary data	When high quality and relevant secondary data is available, statistical techniques such as propensity score matching and instrumental variables can be used to construct a well-matched comparison group.	Can be used whenever there are quantitative outcome and impact indicators that the secondary data also measures. Important to assess the adequacy of the secondary data in terms of sample coverage, when it was collected, and how the indicators were measured. Often secondary data is not available to control for alternative explanations of the observed changes (such as special characteristics of the project

		group making them more likely to succeed).
d. Creative identification of comparison groups	Selecting the best available comparison group on which data can be collected within tight budget and time frames.	An evaluation was conducted in Uganda to assess the impacts of a USAID HIV/AIDS program operating in 16 districts when only 3 weeks was authorized for data collection for the evaluation. It was decided to use the next closest district where the HIV/AIDS program was not operating as the comparison group. As often happens, it was found that there were no differences in outcome between the project and comparison groups, because a larger HIV/AIDS program was operating in many of the comparison areas.
e. Comparison with other programs	Comparison with other programs that use different approaches for achieving similar objectives	CARE Zambia used this approach to evaluate a school program
f. Comparing different types of intervention	This may compare project sites that offer different levels of treatment (“dosage”) or different methods of delivery.	Sometimes these variations were built into the program design, in other cases they occur because of unforeseen circumstances or because each project is given considerable autonomy in terms of how programs are organized.
g. Cohort analysis	Cohorts are similar groups who enter the project at different points in time (for example, students entering sixth grade at the start of each school year). Cohorts can be compared before and after treatments have begun.	In the BMGF project in Zambia several annual cohorts of people were each interviewed at different points in time and the question was asked “How many people have additional sources of income?” Respondents were asked to compare their income before, during and after the project. A comparison was made between the experience of different cohorts and between participating cohorts and between cohort members and the fourth family to the north (not benefiting from the project but close enough to have similar characteristics).
h. Finding a comparison when the program has	Evaluators have to take advantage of natural variations, pipeline-type delays in implementation, or intended or unintended variations in services to	In a state-wide education evaluation in Chennai, India, that covered all schools, useful comparisons were made between different dosages of the treatments, and between outcomes for pilot projects and when the program was rolled out to all schools

universal coverage	reconstruct a comparison group.	in the state. One interesting determinant of outcomes was differences in motivation of participants to succeed.
i. Comparison to a set of similar countries	The effect of national-level interventions can be compared either to a small number of similar countries (usually in the same region) or to a large number of countries using secondary data sets such as MDG indicators, Demographic and Health Surveys, and rating scales such as the corruption scales developed by Transparency International.	Works well when assessing easily obtainable indicators such as life expectancy and mortality rates, exposure to infectious diseases or school enrolment rates. Harder, but possible to use for assessing qualitative changes such as levels of corruption, political participation, gender equality.
j. Citizen report cards	Surveys are conducted in a particular city with a large sample of households who report on their experiences with public service agencies (such as water, electricity, and police). Respondents report on how they were treated when they had a problem to resolve, whether they had to pay a bribe and whether their problem was resolved. The surveys produce baseline performance indicators for each agency which can be used to measure improvements in a follow-up survey or as a reference point for assessing change in other cities.	Can be applied in any urban area where respondents all have access to the same public service agencies. Can be used for a general assessment of public sector service delivery or for a more in-depth look at a particular sector such as public hospitals. It is important that the survey be conducted by an experienced, respected and independent research agency as the findings will often be challenged by agencies that are badly rated. When used for comparison with other cities it is important to ensure that the same methodology and questions are used in each case.
k. Public Expenditure Tracking Studies (PETS)	PETS track in great detail how funds flow from central ministries to the front-line agencies such as schools and health clinics. The proportion of funds reaching front-line agencies and the time taken to deliver funds are calculated. The scores can be used as a baseline for measuring improvements in a follow-up study or to define a counterfactual by using the baseline data collected in one (or ideally several) other cities or countries as the comparison	Can be used to produce baseline data for the delivery of funds to any agency that provides a clearly defined service to the public through a large number of front-line units such as schools, clinics, public transport. The PETS analysis works best where all services are provided through a uniform set of units such as schools and where there is a uniform funding mechanism. More difficult to use for health as there are a wide range of different service providers each using different financial mechanisms.

	group,	
B-3 Qualitatively oriented approaches		
a. Concept Mapping	Interviews or group discussions with stakeholders and experts are used to define a set of performance indicators. Statistical techniques are used to sort the indicators, for example measuring similarities or differences between indicators. The same or a different group will then rate each program or country on each indicator. Ideally the ratings will be conducted at the start and end of the program to measure change, but it is also possible to use concept mapping for retrospective evaluations (where experts rate the amount of change in each indicator over the life of the project).	The value as a counterfactual is enhanced if the sample includes sectors or countries where the program is not being implemented. Can be used to evaluate cross-cutting programs, including those that operate in a number of different countries (for example, gender mainstreaming or community empowerment). In some cases the whole evaluation is based on the ratings produced by experts, while in other cases concept mapping is used to select a set of representative programs or countries for preparing in-depth case studies.
b. Creative uses of secondary data	The object is to find secondary data that describes in sufficient detail what change occurred in reasonably comparable communities during the time-frame of the program. Though this can be done qualitatively, see quantitative section for more details.	Though the particular intervention was not implemented in the comparison communities by the agency being evaluated, one should ascertain what else was going on in the comparison community (internally or externally) that could have contributed to the changes there. [Remember that a counterfactual isn't usually <i>nothing</i> (i.e. no intervention or influence of any kind), it is often <i>something else</i> (though provided by some other source.)]
c. PRA and other participatory group techniques [see B-1 (g. and h) for a discussion of	Community residents or other intended project beneficiaries are organized into groups that conduct exercises using graphical representations to describe the current resources, constraints and problems of the communities (social maps) or to	Frequently used to assess the impacts of rural development programs, often with communities with low levels of literacy (hence the extensive use of graphical and non-verbal techniques.) Also used extensively to assess programs for low-income urban populations. In order to obtain a counterfactual, similar PRA exercises could be conducted in comparable

temporal and relational PRA techniques.]	reconstruct key events in the history of the community or to construct diagrams assessing the impacts of project interventions and other events. Impacts can also be assessed by constructing baseline indicators and then repeating the PRA exercises at the end of the project to assess changes.	communities, asking them to identify changes that occurred in their lives relative to the kinds of changes the program was promoting in the target communities.
d. Process tracing	See B-1 (e)	
e. Compiling a “book of causes”	See B-1 (f)	
f. Forensic methods	See B-1 (j)	
i. Natural variations	See B-2 (b)	
j. comparison with other projects	See B-2 (e)	
k. comparisons among project locations with different combinations and levels of treatment	See B-2 (f)	