FOOD AND NUTRITION TECHNICAL ASSISTANCE

USAID
FROM THE AMERICAN PEOPLE

Technical Note No. 11
March 2006

technicalnotes

Evaluating Title II Development-oriented Multi-Year Assistance Projects

Gilles Bergeron, Anne Swindale,
Megan Deitchler and Paula Bilinsky

Evaluating a Title II development-oriented multi-year assistance project (MYAP) involves assessing its outcomes and impacts, that is, verifying the extent to which project activities are associated with intended changes in the practices and well-being of the beneficiary population.[1] Evaluation objectives may range from simply measuring the level of change in indicators of well-being, to attributing a change in the level of those indicators to the intervention being implemented. The focus of this Technical Note is to lay out the various evaluation design options open to Title II project managers. Choosing among designs depends on the answer to one basic question:

How confident do project managers need to be that the changes they see are a result of project activities?

The answer to this question leads to the selection of one of four types of evaluation designs applied at the beginning or baseline of a project (Pre) and at the end or final evaluation of the project (Post) [2]:

- Simple Pre-Post (Type I);
- Pre-Post with Control Groups (Type II);
- Pre-Post with Treatment of Determinants and Known Confounding factors (Type III); and
- Pre-Post with Control Groups and Treatment of Determinants and Known Confounding factors (Type IV) (see Table 1):

Table 1: Types of designs	**No Control Groups included**	**Control Groups included**
Determinants and Confounding Factors not considered	Type I **(Adequacy)**	Type II **(Plausibility)**
Determinants and Confounding Factors considered	Type III **(Plausibility)**	Type IV **(Plausibility)**

[1] This stands in contrast to the *monitoring* of Title II projects, which chiefly meant to ensure that inputs, processes and outputs are implemented as planned.

[2] Terms often used in evaluation design are adequacy (Simple Pre-Post), and plausibility (Pre-Post with Controls, and Pre-Post with Treatment of Determinants and Known Confounding Factors, with or without Controls) designs. Probability designs (referring to randomized trials) are not examined here, as this type of design is not appropriate in the context of routine Title II project evaluations (see Habicht and Victora, 1999).

[3] In certain situations, it may be safe to assume that no other factors came into play: for instance, immunization coverage can be safely attributed to an immunization campaign, if it is known that no other organizations have been vaccinating children in the area; and a Type I evaluation will provide project managers all the assurance they need to state that post-campaign immunization coverage is indeed a result of their intervention.

Adequacy Design: Simple Pre-Post Evaluation (Type I)

A simple pre-post evaluation design answers the question of whether a significant change has occurred in an outcome/impact indicator by measuring and comparing its level before (at baseline) and after the intervention (at final evaluation). Simple pre-post designs are the simplest and least expensive of all the designs described here—chiefly because they do not require that control groups be included. This absence of control groups, however, is also the key weakness of this type of design, as it does not allow the analyst to separate out the outcome/impact of the project from those of other possible factors.[3] Change in the indicator could well have occurred even in the absence of the project, due to events happening at the household, national or other levels unrelated to project activities.

	Intervention area population
Baseline	
Final Evaluation	

Attribution of effects to project activities can therefore not be made with a Simple Pre-Post design. All that may be said is that the beneficiary population now exhibits the levels shown in the outcome/impact indicator at the final evaluation stage. Also, if no change is observed, the Simple Pre-Post design provides no guidance as to whether the failure is due to project implementation issues or whether the project activities averted what would have been a greater decline in the value of the indicator.

USAID's Food for Peace Office does not require that evaluations attribute effects to the project. Thus there are no compelling reasons preventing a project from selecting a Simple Pre-Post design and in many cases, this type of design is appropriate for a Title II project. If Type I design is used though, it will be important that the other elements of the general M&E system provide sufficient information on the inputs, processes, and outputs to lend credibility to the claim of association between activities and effects. However a project manager may wish to state that the association is not only credible, but plausible. If so, a more powerful evaluation design is required.

Plausibility Designs (Types II, III and IV)

"Plausibility designs" refer to a family of designs that vary in complexity from the simple comparison of the beneficiary population to a control group before and after the intervention, to designs that compare those same groups before and after the intervention, while simultaneously controlling statistically for other determinants and confounding factors. Accordingly, attribution statements made from such designs go from weak to strong: at the weakest level a simple comparison between intervention beneficiaries and a control group is used to try to separate the influence of unaccounted-for factors (Type II). At the strongest level (Type IV), most alternative explanations have been identified, measured and accounted for.

Type II: Pre Post Design with Control Groups: This most basic of all plausibility designs is illustrated by the use of a 2X2 comparison table (see below). This design requires that baseline and final evaluation data be collected for the same indicators in both the intervention area population and in the control group. The attributable influence of the project is measured by quantifying the difference in levels of the outcome/impact indicator between the two groups at the two points in time. Specifically (and provided the intervention is successful) the outcome/impact indicator should:

(i) be at the same level in the two groups at the start of the intervention;

(ii) indicate that a change has taken place in the beneficiary population following the intervention;

(iii) indicate that the level of change in the control group is smaller than in the intervention group;

(iv) indicate that the difference in outcomes/impacts between the population in the intervention area and the control group after the intervention is statistically significant.

	Intervention area population	Control group
Baseline		
Final Evaluation		

Verifying that those four conditions apply makes it reasonably plausible to associate a change in outcomes/impacts in the beneficiary population with the intervention. However, doubts may remain that other factors could explain that outcome/impact. In a project that intends to raise agricultural output, for instance, what if the control group was subject to different climatic conditions than the beneficiary population during the period of intervention? This could explain a difference in outcome/impact. And here indeed lies the main problem in using control groups: individuals or groups that did not benefit from the intervention may differ from project beneficiaries in significant ways. To address this, a second type of plausibility design may be used.

Type III: Pre-Post Design with Statistical Treatment of Determinants and Known Confounding Factors: As defined in Box 1, "determinants" are any features that predictably influence the outcome/impact of the intervention—so much so, in fact, that projects often chose to attack a problem by addressing its determinants rather than its manifestations. Classic examples of determinants are, for projects that aim at improving child nutritional status: the child's health, infant and young child feeding practices, or household food access. Or for agricultural projects: farmer technical know-how, seed types, and soil fertility. Known confounding factors, by contrast, are factors that can influence the outcome/impact, but over which the project has no control—weather is a good example here (see Box 1). Once measured and quantified, determinants and known confounding factors can be accounted for ("kept constant") using multivariate statistical analysis techniques.

This type of plausibility design does not involve a control group but requires that information be collected at baseline and final evaluation on both the outcome/impact, its determinants and on known confounding factors. In that regard, it is best contrasted to the Simple Pre-Post Design: in a nutrition intervention, for example, a Simple Pre-Post Design would collect data on the nutritional

Box 1: Definitions

Control groups are similar in key respect to the intervention populations, except in their exposure to the intervention.

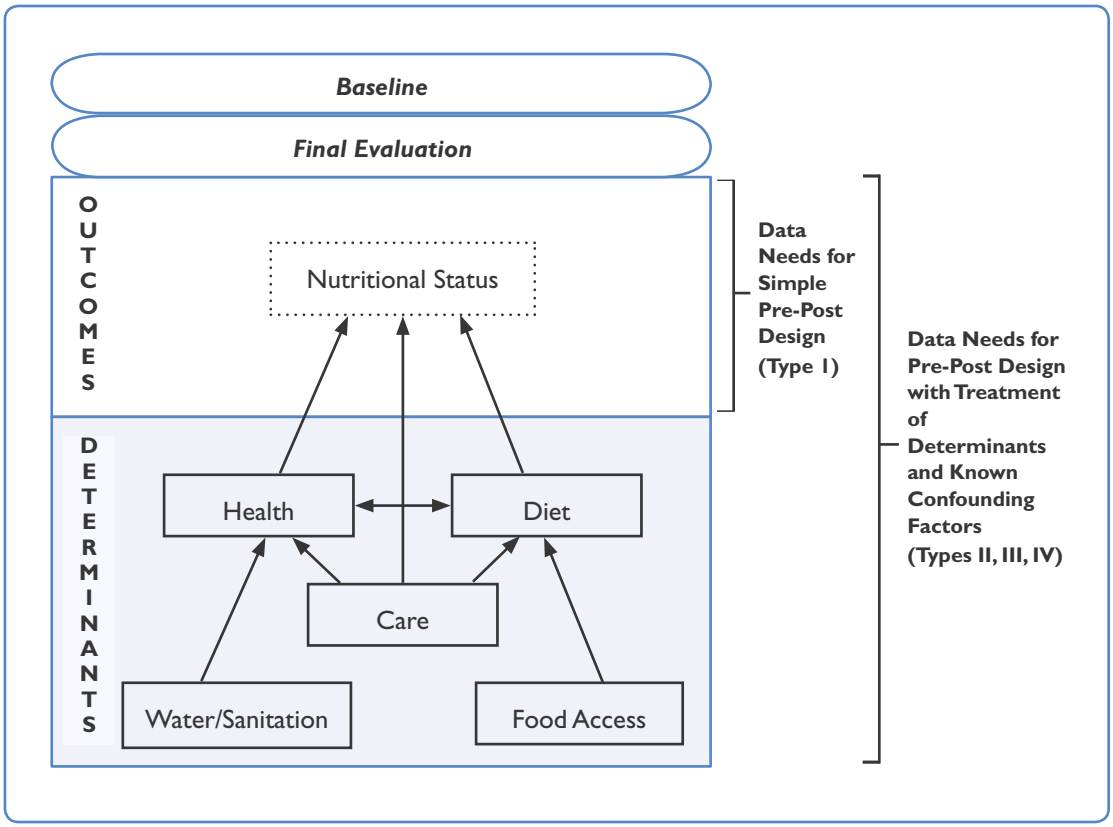
Determinants are features known to predictably influence the effect of an intervention (and as such, can be themselves subjected to intervention). Also, being known and measured they can be statistically kept constant in the analysis.

Confounding factors are any systematic feature intrinsic to a group that may affect its comparability to the other group. Two main concepts (*confounders* and *effect modifiers*) are covered under our use of the term "confounding factors." A ***confounder*** is an alternative explanation for an observed outcome/impact; whereas an ***effect modifier*** is a factor that influences the outcome/impact of the intervention. For instance, favorable rainfall (weather) could be an alternative explanation for a successful agricultural outcome. By contrast, ethnicity or education level may modify the effect of the intervention-say if a particular ethnic group responds more to the intervention than another, or if mothers with greater education are more responsive to health messages. Like determinants, known confounding factors can be measured and statistically kept constant; whereas unknown confounding factors (which by definition cannot be defined nor measured) can only be accounted for by the use of a control group (as it is assumed that the random sampling of beneficiaries and controls ensures a similar distribution of such confounding factors among the two comparison groups).

status of the child at baseline and final evaluation. By contrast, a Pre-Post Design with Statistical Treatment of Determinants and Known Confounding factors would collect data, not only on nutritional status, but also on the determinants of malnutrition–which could include, for example, the health, care and diet of that child; and on known confounding factors (ethnicity, mother's education, etc) (see Figure 1—known confounding factors not shown).

Choosing Type III design will have important implications on the cost of the surveys. This is examined in the next section "Choosing a Design Type."

Figure 1. Pre-Post Design with Statistical Treatment of Determinants and Known Confounding Factors



Type IV: Pre-Post Design with Controls and Statistical Treatment of Determinants and Known Confounding factors: As its name implies, this design type combines both the use of controls and the statistical treatment of determinants and known confounding factors: it keeps those constant, while controlling also for unknown or unmeasured confounding factors by the use of a control group. It is the most powerful of all the design types presented here: assuming success in the intervention, it will provide the information needed to say:

(i) that the severity of the problem was reduced in intervention areas;

(ii) that change in the known confounding factors do not explain the observed improvement

(iii) that changes in unknown confounding factors should not explain the improvement

(iv) that the severity of the problem did not improve (or improved significantly less) in areas without the intervention[4];

(v) that there is a demonstrable association between the degree of exposure to the intervention and the size of the improvement.

This combination of facts would greatly reduce the possibility that a change in the outcome/impact indicator in the intervention population is due to causes other than the project intervention. This level of certainty would come at a high cost, however, as it would involve using an extensive interview instrument—to capture information on determinants and known confounding factors—as well as a doubling of the sample size—to accommodate the inclusion of a control group.

The identification of **determinants** is important, as it usually leads the way to intervention design. It is best done using conceptual frameworks, such as "UNICEF framework." According to this framework, the immediate or proximate determinants of malnutrition are inadequate care, diet and illness. The underlying factors of poor care, diet and illnesses are 1) inadequate household access to food, 2) inadequate environment, including poor water and sanitation, and 3) Insufficient health services. Having such a model in hand is useful when it comes to identify determinants since access to food, water and sanitation and care knowledge are now related to the impact (nutritional status).

[4] Note that these first three statements would also apply to the previous design, "Pre-post design with Statistical Treatment of Determinants and Known Confounding Factors." The critical difference is in the fourth and fifth statement which only the use of a control group will allow for.

Choosing a Design Type

The key features of the four design types presented so far are summarized in the table here.

Type of design	What it does	What it says	When might be considered	What data it needs
Type I: Simple Pre-Post	Compares data on outcomes/impacts in intervention population before and after the intervention	Determines whether a change has occurred at the level of the outcome indicator	No competing explanations, or not able to include control groups and/or do not know confounding factors	Data on outcomes/ impacts collected among the intervention population
Type II: Pre-post with Controls	Compares outcomes/impacts across intervention population and control groups before and after the intervention	Plausibly associates a change in outcome /impact levels to the intervention	Suspect competing explanations, and is able to account for confounding factors by including control groups	Outcomes/impacts data collected among the intervention population and control groups
Type III: Pre-post with Statistical Treatment of Determinants and Known Confounding Factors	Compares outcomes/impacts in intervention population before and after the implementation, controlling for determinants and known confounding factors	Plausibly associates a change in outcome/impact levels to the intervention and quantifies the role that other known factors may have played in affecting it	Suspect competing explanations, and is able to obtain information on confounding factors, but is unable to include control groups	Data on outcomes/ impacts, determinants, and known confounding factors collected before and after the intervention
Type IV: Pre-post with Controls and Statistical Treatment of Determinants and Confounding Factors	Compares outcomes/impacts across the intervention population and control groups before and after the intervention, controlling for determinants and confounding factors	Plausibly associates a change in outcome/impact levels to the intervention and quantifies the role that other known confounding factors may have played in affecting it, while controlling for the outcome/impact of potential unknown factors	Suspect competing explanations, and is able to obtain information on confounding factors, and is able to include control groups	Data on outcomes/ impacts, determinants, and known confounding factors collected among intervention population and control groups before and after the implementation

Once the options for selecting a design type are clarified, choosing among them should proceed based on the objective of the evaluation, the audience for the evaluation, costs, and the capacity to obtain and use the information needed.

- *Objective of evaluation*: Summary program information to indicate that the intervention implementation is making adequate progress towards the outcome may not require a complex evaluation design. Determining what affects the effectiveness of a program, however, requires that other determinants or confounding factors be kept constant, thus requiring the use of Types II, III or IV Types.
- *Audience:* who uses the information, and how, affects the level of detail needed and the certainty of the evaluation findings produced. Scientists and project managers often want more details, greater precision, and want to understand how things relate to one another. Funding agencies on the other hand may be more interested in confirming the usefulness of their investment and will be satisfied if the project shows positive outcomes/impacts. How the information is used also matters: strategic considerations, the testing of a new approach, a new country situation, or future funding needs, may all give a project manager good reasons to require greater details and better understanding of the relationship between the intervention and documented improvement.
- *Costs:* three key aspects will affect costs: sample size, length of interview and analytical needs:
    - Sample size needs are the smallest for the Simple Pre-Post and for the Pre-Post with Statistical Treatment of Determinants and Known Confounding factor (Types I and III) designs. The other two will likely require a doubling of the sample size to accommodate the inclusion of control groups.
    - Questionnaires for Types III and IV require more extensive information than Types I and II, to document the determinants and known confounding factors. This will affect the length of the interviews in the field, and the complexity of data collection and entry with proportional increases in field costs.
    - As the design becomes more elaborate, analytical requirements become more complex, demanding higher investments in analysis and write up.
- *Capacity:* A critical question the project manager must answer before choosing a particular design type is whether the human and technical resources needed to carry out that analysis are available—indeed it would serve little purpose to implement a complex, costly evaluation to find later that the project does not have the technical resources to properly clean, analyze and report on the data.

Conclusion

An effective evaluation strategy begins with well-defined objectives: Title II project managers should know from the start what questions they want to see answered in their final evaluation. This prior knowledge is essential in order to select a design that fits the task. Having a right design type, in turn, is a critical guide in deciding what data to collect: evaluators frequently overload their survey instruments with questions that provide little added value, while taking precious time away from respondents and spending unnecessary resources. In a well designed evaluation every piece of data finds its rightful place and no superfluous information is collected. Schedules, workloads, logistics and budgets also benefit from the long-term perspective granted by a carefully planned evaluation design. Last but not least, no matter what evaluation design type is selected, all should be conducted in an equally rigorous manner. The differences between design types are not in how "easy" they are, but in the purposes for which they are conducted.

Reference: Habicht, J.P., C.G. Victora, and J.P. Vaughan, 1999, "Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact," *International Journal of Epidemiology* 28:10-18.

Related Reading

FANTA Technical Note No. 10, *Monitoring and Evaluation Framework for Title II Development-oriented Projects*

Additional titles in the Technical Note Series are available at the FANTA web site:

www.fantaproject.org/publications/

Food and Nutrition
Technical Assistance Project
Academy for Educational Development
1825 Connecticut Ave., NW
Washington, DC 20009-5721
Tel: 202-884-8000
Fax: 202-884-8432
E-mail: fanta@aed.org
http://www.fantaproject.org

This publication was made possible through the support provided to the Food and Nutrition Technical Assistance (FANTA) Project by the U.S. Agency for International Development's (USAID) Office of Food for Peace of the Bureau for Democracy, Conflict, and Humanitarian Assistance and the Office of Health, Infectious Disease and Nutrition of the Bureau for Global Health, under terms of agreement No. HRN-A-00-98-00046-00 awarded to the Academy for Educational Development (AED). The opinions expressed herein are those of the authors and do not necessarily reflect the views of USAID.

Recommended citation: Bergeron, Gilles, Anne Swindale, Megan Deitchler and Paula Bilinsky. *Evaluating Title II Development-oriented Multi-Year Assistance Projects*. Washington, D.C.: Food and Nutrition Technical Assistance Project, Academy for Educational Development, 2006.