Economic efficiency studies typically have two quite distinct components. The first is a program cost study, which typically involves both primary data collection and the assembly of data from program accounting records. The most important costs are typically the direct cost of the service or benefit and the various administrative costs of delivering the service. Service delivery costs usually occur at State and local levels of program operation as well as at the Federal level. Even in programs in which the Federal Government makes a payment for administration, State or local operators often incur costs beyond those reimbursed. In addition to these costs, many evaluations must consider costs to participants, most commonly time expended in complying with the requirements for program participation, but sometimes also tax payments or work expenses associated with income received. Some participant costs can be difficult to express in monetary terms, such as the potential for job loss associated with taking off time from work or negative psychological consequences of receiving assistance. When other stakeholders are involved in service delivery, as food retailers are in redeeming food stamps and WIC vouchers, costs to these groups may have to be measured as well.

The second major component of economic efficiency studies consists of transforming the impact estimates and program costs into comparable time periods and perspectives. Often the program costs for a particular participant are incurred quickly, during a brief period of program participation, while impacts develop slowly and endure for some years. Efficiency studies are therefore typically framed in terms of the "participation lifetime" (i.e., all of the costs and impacts that are incurred between the time the participant comes in contact with the program and the time when impacts cease to be counted). The studies usually recognize explicitly that one party's cost may be another party's benefit. Thus, cost and effect data are typically presented from at least three perspectives: that of the taxpayer, that of the participant, and that of society as a whole (usually conceived as the net of all parties' perspectives).

When costs and benefits are naturally measured and expressed in dollar terms, it is easy and meaningful to calculate a benefit/cost ratio or net benefit per participant. When translating effects into monetary units requires heroic or tenuous assumptions, however, it is seldom useful to make the translation. This is most often the case with food assistance and nutrition programs, whose nutrition and health impacts are not usually measured in dollar terms.[23] Even when some effects or costs cannot reasonably be monetized, however, the efficiency study is a critical requirement for policymaking. Only when program costs and effects are presented together can the policymaker understand what the program returns for a dollar spent.

Other Program Evaluation Situations

Most evaluations of USDA's food assistance and nutrition programs will probably be overall evaluations of the ongoing programs or demonstration interventions. Two other evaluation situations, which arise less frequently, are discussed in this section. In one situation, the evaluation concerns a change to an ongoing program that is implemented at the same time in all program locations rather than being introduced as a pilot or demonstration initiative. This situation is distinguished by a very limited set of options for evaluation design. In the second situation, the evaluation focuses on a single component of an ongoing program, attempting to distinguish its impact within the overall program package.

Impact Evaluation of Programwide Modifications to Ongoing Programs

Major national programs sometimes undergo important general changes, such as in eligibility criteria or the nature of program benefits or services. Such changes often result from legislation requiring nationwide implementation of the change on a particular date. Unlike the demonstration trial of a program modification, this situation offers no opportunity to observe the old rules and new rules operating in parallel for different individuals or areas.

A current example is the PRWORA, which radically changed the way participating family child care homes are to be qualified for eligibility for cash subsidies in the CACFP. Prior to PRWORA, a fixed per meal subsidy was paid to all participating family child care homes for all children who were served meals in the

[23]There are exceptions, such as the study in which Devaney and colleagues (1991) calculated that the savings in Medicaid expenditures achieved by raising the average birthweight of newborns more than offset the costs of running the Medicaid program. (Devaney's is not a full cost-benefit study, however, because only some costs and some benefits were considered.)

homes. PRWORA allows full subsidies only for homes located in low-income areas or operated by low-income providers. Other providers receive reduced subsidies, except for individual children who meet a means test.

The law mandated that the changes take effect in July 1997. Because of concerns raised about the possible effect of these changes on children and child care providers, the legislation also mandated an evaluation of the impact. The evaluation mandate was not accompanied by permission to phase in the changes or otherwise to operate the new and old systems in parallel.

Such legislatively mandated, programwide reforms have generally been interpreted as precluding randomized experimentation. Quasi-experimental designs are therefore employed, as described below.

Quasi-Experiment 8 Pre-Change Participants vs. Post-Change Participants

Features:

Impact estimate: Difference between outcomes for program participants before the change and program participants after the change.

Key requirement: Obtaining pre-change outcome measures.

Advantage: The only option available in some cases.

Disadvantage: Unreliable because of confounding with contemporaneous factors.

The Three Questions:

Alike before exposure? Reasonably similar.

Difference solely from intervention? No, other events in same time frame may cause differences.

Full force of intervention represented? Yes, if change is implemented fully.

Quasi-Experiment 8: Comparing Pre-Change Participants to Post-Change Participants

This is one of the simplest and weakest of quasi-experimental designs. Outcome data are collected for all or a sample of program participants before the change is implemented. After the change is implemented, the data collection is repeated, again for all or a representative sample of program participants. Regression adjustment is used in estimating impacts to account for any shifts in the measured characteristics of the participant population that may occur between the pre- and post-change periods (unless the intervention itself is expected to cause such shifts) (see box).

The evaluator's main challenge in this situation is usually to obtain appropriate outcome measures during the pre-change period. Legislative changes often must be implemented quickly, and if the changes have not been anticipated, there may be insufficient time to mount a primary data collection effort. This may require the evaluation to rely on administrative data, one of the large periodic national surveys, or a previous study.[24] In one interesting example, the expectation of welfare reform legislation led to what might be called speculative data collection. Research planning to measure pre-welfare reform outcomes for families on AFDC had to begin several years before PRWORA was passed in 1996—with no firm knowledge about the nature of the changes to be enacted and based only on the firm belief that some sort of welfare reform would be legislated within the next few years (Rossi, 1999).

Even when the data collection challenge is met, this design is very weak. Important national events occurring in the same time period as the program changes may influence the outcomes of interest. The CACFP changes provide a good example. Implementation of these changes coincided with welfare reform and an unprecedentedly strong labor market, both of which are expected to have great influence on the demand for and supply of child care. The pre/post evaluation of the CACFP changes, no matter how carefully designed, will not be able to determine how these major changes influenced the measured difference in outcomes.

[24]In the CACFP example discussed above, pre-change measures of the characteristics of meals served will come from a study conducted several years previously (Fox et al.,1997).

Time Series Analysis

Time series analysis (Quasi-Experiment 4) is an alternative to the simple pre/post comparison in evaluating the impact of programwide changes. It is particularly attractive when the outcomes of interest are measured in data series generated by the program itself. Most examples of this approach, therefore, deal with the impact of program changes on participation. Quite sophisticated time series modeling has been used to estimate the impact of eliminating the purchase requirement in the FSP and the impact of the 1981 eligibility changes in AFDC (Moffitt, 1986). When the outcomes of interest concern nutrition and health status, however, adequate data series are rare.

An important limitation of the time series approach, even when appropriate outcomes are routinely measured, is the need for multiple observations in the post-change as well as the pre-change period. Virtually all data series contain period-to-period fluctuations that are not part of the general trend. When a deviation from the trend line occurs in the last one or two periods of the series, time series models cannot readily determine whether it represents a lasting change from the trend line or a temporary fluctuation. Moreover, if multiple events occur in the same period, such as welfare reform and the CACFP changes, time series analysis is no more effective than simple pre/post analysis, as neither can distinguish the effects of the different events.

Impact Evaluation of Ongoing Program Components

Most policy decisions about ongoing programs are not "go/no go" decisions about the program as a whole, but decisions about whether to modify or eliminate particular program components. A program component of interest can be an element of the service package, such as nutrition education in WIC or employment and training in the FSP. Alternatively, the program component may be a portion of the service delivery mechanism, such as food stamp cashout or electronic benefit transfers.

Many program components offer the advantage (from an evaluation perspective) of not being tightly specified in the authorizing legislation or program regulations. In these situations, State or local operating entities can create variations in the program by virtue of the ways they choose to implement the program component (see box). WIC nutrition education, for example, varies considerably in the frequency with which nutrition education sessions are offered, the topics covered, and the format in which the service is delivered.

Natural or Planned Variation Studies

Features:

Impact estimate: Difference between outcomes for program participants with alternative levels or styles of the intervention.

Key requirement: Numerous representations of each version of the intervention.

Advantage: Can identify "good practices" for replication.

Disadvantage: No estimate of overall impact of the intervention.

The Three Questions:

Alike before exposure? No, but some differences can be taken into account.

Difference solely from intervention? No. Other location-related forces may cause differences.

Full force of intervention represented? Only the difference in interventions is represented.

Natural Variation Studies

The natural variation evaluation, which is applicable in the situation described above, can be considered a "dose response" study. The underlying proposition is that more of the intervention (or more of a particular quality of the intervention) leads to more of the outcome. The design does not yield an estimate of the impact of the intervention itself.[25]

[25]The inference is sometimes drawn that, if a higher level of the intervention has a significant impact relative to a lower level, it also has a significant impact relative to no intervention at all. This inference may be incorrect if, in the absence of the intervention, the participant would have done something different that would have resulted in equivalent or better outcomes.

Because differences in program features are normally introduced by different operating units, these units become key elements of the design. In fact, the design does not literally compare differences in the program; rather, it compares different sites that happen to implement the program differently. Thus, the design is highly vulnerable to confounding the effect of program variations with the effect of other factors that differ among sites, which may range from regional economic characteristics to the abilities of site staff. The only way to reduce this vulnerability is to have multiple sites representing each programmatic variant.

Implementing this design requires first defining meaningful variants of the program component and then identifying a number of sites that implement each variant. Outcomes are then measured for all or a sample of participants in each group of sites. Participant outcomes are modeled as a function of the program variant they face, their site, and an array of participant characteristics.

Planned Variation Studies

To the extent that program legislation and regulations allow program operators discretion in shaping program components, they also open the possibility for planned variation. In a planned variation design, the agency sponsoring the evaluation (or sometimes the evaluator) arranges for the use of specified variants of the program component by particular sites or in particular circumstances.

If planned variation is feasible, a randomized experiment is likely to be possible and is the preferred design. Individuals or aggregates of individuals are randomly assigned among the variants being tested. Differences in outcomes can be attributed to the differences in the program component. If randomized experimentation is precluded, the possibilities include the same array of designs described earlier for evaluating demonstration modifications to ongoing programs.

Parting Words

This report has noted, at several points, that randomized experimentation is the preferred design for impact evaluation in practically all situations. However, the bulk of the discussion has been devoted to the many quasi-experimental designs that are often used in place of randomized experimentation.

Lest the word count distort the message, we must reemphasize here the importance of exerting all possible efforts to use randomized experiments. For programs that deliver services and benefits directly to individuals and families, randomized experimentation is the only design that, properly applied, is guaranteed to produce unbiased estimates of program impact. All other designs are vulnerable to some bias. Their sources of bias can sometimes be described, but the direction and magnitude of the bias cannot be measured reliably. Thus, all the nonexperimental designs have some substantial probability of producing answers that are far from the truth—which can lead to inappropriate policy decisions that may affect millions of people and billions of dollars of public expenditure.