Effects of Food Assistance and Nutrition Programs on Nutrition and Health
Volume I, Research Design

William L. Hamilton and Peter H. Rossi

Introduction

This report reviews the research designs available to evaluators as they contemplate assessing the impact of USDA's food assistance and nutrition programs on nutrition and health outcomes. The focus of the discussion and the examples presented concern the five main food assistance and nutrition programs: the Food Stamp Program (FSP); the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC); the National School Lunch Program (NSLP);, the School Breakfast Program (SBP); and the Child and Adult Care Food Program (CACFP). The research designs discussed cover the gamut of those commonly mentioned in the evaluation literature, and are applicable not only to other Federal food assistance and nutrition programs but to practically all social programs that directly serve individuals and families.

In the same vein, although the discussion refers most often to impact on participants' nutrition and health outcomes, the designs are generally applicable with any outcomes measured for individuals. However, the availability or nonavailability of nutrition and health outcome data in existing data sets and the procedures required to collect these data sometimes constrain the set of design choices for evaluating food assistance and nutrition programs.

The impact of a program or other intervention is defined as the difference between what happens in the presence of the intervention and what would have happened in its absence. What would have happened in the program's absence is generally called the "Counterfactual."

All of the evaluation designs discussed in this report involve measuring outcomes that occur in the presence of the intervention and comparing them to some representation of outcomes in the Counterfactual.[1] The main difference in the designs lies in the ways they represent the Counterfactual.

Establishing the Counterfactual—estimating what would have happened without the program—is usually accomplished by examining a population that has not been subjected to the intervention being evaluated. What makes the task difficult is the fact that people[2] who become participants in a social program are often quite different from those who do not, because they either have been selected for participation or have selected themselves (Campbell and Stanley, 1963). These selective processes may make participants different in important ways from those who do not participate. These differences include not only people's permanent characteristics, such as their gender or race, but also transitory characteristics such as their current income or employment, the opportunities they face, and the experiences they have had. Many of the transitory characteristics result from the time and place in which people live, which means that similar people in a different time or place may not appropriately represent the Counterfactual. All these influences may contribute to selection bias, which distorts the evaluation of a program's impact.

[1]This excludes prospective simulation approaches to evaluation. In these approaches, outcomes in the presence of the intervention are not measured directly, but projected on the basis of prior information or assumptions about the effects of the intervention or its elements.

[2]Evaluation designs often focus on units other than people, either aggregations of people (e.g., families, students in a school, the population of a county) or operating entities (program offices, schools, businesses). For simplicity of presentation, the discussion generally refers to individuals rather than aggregations or other entities.

The general strength of an evaluation design in a particular situation can be assessed through the following three questions:

- Is the population representing the Counterfactual equivalent in all pertinent respects to the program population before that population is exposed to the intervention?
- Is the intervention the only force that could cause systematic differences between the two populations once exposure begins?
- Is the full force of the intervention applied to the program population, and is none applied to the Counterfactual population?

The technically preferable evaluation design in any situation is one that provides strong affirmatives to all three questions. In the sections that follow, these three questions will be used to characterize the conceptual strengths and weaknesses of each design.

The report has three main sections. The first section deals with designs for evaluating ongoing national programs, such as the five major food assistance and nutrition programs. Because these programs are available to practically all potentially eligible people nationwide, and because they have been operating for a long time, they pose particularly difficult challenges for evaluation.

The second section focuses on designs applicable to evaluations of demonstration initiatives that would modify existing programs or create new ones. Many food assistance and nutrition program evaluations are likely to fall into this category, which fortunately tends to be more tractable. The third and final section of the report considers two less common evaluation situations: evaluation of a mandated programwide reform and natural- and planned-variation evaluations of program components.

Impact Evaluation of Ongoing Programs

The question of whether and how much the major food assistance and nutrition programs affect the nutrition and health outcomes of participants has obvious policy importance. These programs account for very sizable Federal expenditures—$33.5 billion in fiscal year 1998—but little scientifically sound evidence exists on the programs' impacts, particularly their effect on nutrition and health outcomes.

The ongoing food assistance and nutrition programs have two characteristics that make it extremely difficult to assess their overall impact on participants' nutrition and health outcomes. First, they are essentially universally available throughout the United States. For practical purposes, there exists no current population that has not been exposed to the programs, where people are considered "exposed" if they have reasonable access to information about the program and would be able to participate if they applied and were found eligible. Second, the programs have operated nationally at a substantial scale for a minimum of two decades. This means that, even if one could find measures of the relevant outcomes for a period before the programs began, no identifiable population in the preprogram period is likely to have permanent and transitory characteristics equivalent to those of today's participants.

Of the several possible research designs described in this report, only randomized experimentation is actually capable of providing reliable estimates of the programs' impacts. However, randomized experiments have not been applied to measure the overall impact of these programs to date (although they have been used to measure the impact of program modifications), and we recognize the likelihood that such experiments may not happen in the near future. For this reason, we discuss several possible quasi-experimental designs. The quasi-experimental designs, which are second-best