6 Evaluation Methods for Program Entry Effects

Robert Moffitt

Most program interventions are intended to change the characteristics of the recipients on a program, to change the caseload of the program, or both. In the case of welfare programs, most of the major interventions of policy interest over the past twenty years, both those that have been implemented as well as those that have only been proposed or tested, have been aimed at the twin goals of increasing the labor supply and earnings of welfare recipients and of the low-income population in general, and reducing welfare caseloads. However, most of the designs of program evaluations, both those using experimental methods and those using nonexperimental methods, have focused considerably more attention on the former goal than on the latter. Evaluations of a negative income tax, for example, were primarily concerned with the effects of such a program on labor supply. Evaluations of training programs for welfare recipients, to take another example, have as their primary focus the estimation of the effects of such programs on the earnings and employment of trainees. The excessive focus on earnings and labor supply effects has led to a surprising, indeed disturbing, lack of attention to effects on program entry. While considerable attention has been paid to program effects on exit rates from welfare, estimation of the effects of a program intervention on the caseload requires the estimation of effects on the program entry rate as well as the exit rate; both together determine the caseload effect. Entry effects may be particularly important in the long run, for such effects may develop gradually as the eligible population becomes aware of the program change put in place by the intervention.

The evaluation of the effect of an intervention on program entry, or on the caseload as a whole, requires very different methods than an evaluation of its effects on earnings or labor supply. Most evaluation designs for the estimation of earnings and labor supply effects are based on an examination of individuals who are on the welfare rolls; this is natural since the treatments under consideration (training, for example) can only be received by those who are welfare recipients. However, such an approach precludes the estimation of entry effects, for such estimation requires an examination of the extent to which those off the rolls go onto them in greater or lesser numbers than was the case prior to the intervention.

It is also possible that the earnings and labor supply effects estimated with conventional designs are themselves adversely affected by the neglect of attention to program entry. For example, if program entry rates increase as a result of an intervention, and if those who newly enter the rolls have systematically different earnings and labor supply effects than those initially on the rolls, the final average earnings and labor supply effects will be altered. In addition, if a conventional design is used to study the effect of an intervention on program exit rates—by examining the rate at which those on the rolls move off them—the estimates so obtained may also be contaminated by program entry effects. If, again, program entry increases and if those who enter the rolls have different exit rates than those initially on the rolls, final program exit rates will be altered. In both of these cases, proper estimation of earnings, labor supply, and exit-rate effects cannot be obtained in the first place without proper attention to entry-rate effects.

This chapter discusses evaluation methods for the estimation of the effects of interventions on program entry. Methods for the evaluation of such effects pose special difficulties and require attention to several features of design that are quite different from those that usually generate attention. In the following section, the general issues involved are illustrated by a review of several major past evaluations of the impacts of interventions in the Aid to Families with Dependent Children (AFDC) program. The subsequent section discusses possible evaluation designs for the estimation of entry effects. The bearing of these issues on the estimation of earnings effects is the subject of the next section. A summary and conclusions appear in the final section.

Past Evaluations of Welfare Interventions

Over the past twenty years there have been several major evaluations of actual interventions in the AFDC program and of proposed interventions in welfare programs in general. These include (1) one set of experiments of the work-incentive effects of benefit-reduction rates on labor supply, the income maintenance experiments; (2) two sets of evaluations of the major pieces of AFDC legislation prior to 1988, the 1967 Amendments to the Social Security Act and the 1981 Omnibus Budget Reconciliation Act (OBRA); and (3) one set of evaluations of the work and training programs spawned by the OBRA legislation, best exemplified by the experimental evaluations conducted by the Manpower Demonstration Research Corporation (MDRC).[1] In addition, evaluations are currently under way or in the planning stage for the most important recent piece of AFDC legislation, the Family Support Act of 1988.

Despite the high quality of many of these evaluations in numerous respects, not a single one has addressed the issue of entry effects.

Income Maintenance Experiments. The income maintenance experiments were tests of the effect of a negative income tax on labor supply; they were conducted in four different locations during the 1960s and 1970s (for reviews see Moffitt and Kehrer, 1981; SRI International, 1983). The explicit aim of the experiments was to estimate the effect of lowering the benefit-reduction rate on work incentives, although implicitly it was understood that the public goal behind such an intervention was to lower the caseload as well. The experiments were conducted by randomly assigning members of the low-income population in the different areas to experimental and control groups, the former to receive the negative income tax with its lower benefit-reduction rate and the latter to receive only the then-existing set of welfare benefits.

The experimental design reflected a lack of interest not only in entry-rate effects but in participation-rate effects in general. Specifically, in all of the evaluations the members of the experimental group were not permitted not to receive payments to which they were financially entitled. To maintain compliance with the rules of the experiment and to avoid being terminated from the experiment by the program operators for noncompliance, all families were required to submit income report forms on a monthly basis. If the income of the family was sufficiently low to warrant a payment, a check was automatically issued. A family not wishing to participate had only the option of attrition, in which case no further information on its status would be collected by the survey

1. There have been many other evaluations of AFDC and other interventions; these are only those that are best known. See Chapter 1 in this volume by Greenberg and Wiseman for a survey of others following OBRA, and see Moffitt (forthcoming) for a survey of other training evaluations such as Supported Work.

staff. Consequently, it was not possible in the income maintenance experiments to estimate the participation rate in a negative income tax, much less how that participation rate would be decomposed into exit and entry rates.[2]

It is also possible that the experimental labor supply estimates were affected by this defect of the design. In a program instituted on a national level, participation among eligibles would not be required and would certainly be less than 100 percent. Consequently, some of those receiving payments in the experiments, whose labor supply was therefore affected, would probably not participate in a nationally implemented program. If it were the case that such nonparticipants had systematically different responses to the experiment than did those who would participate in a national program—for example, if they had stronger desires to work off the rolls and hence had lower responses to the negative income tax—the average labor supply effect estimated in the experiments would not be a correct estimate of that which would obtain in a national program.

If the experiments had considered designs that would permit the estimation of participation-rate effects, a fundamental dilemma would have appeared which arises in virtually all experimental evaluations, including those discussed below (for example, the welfare employment evaluations conducted by MDRC). If an experiment had been constructed to permit estimation of the participation rate, the randomization would have had to be conducted on the eligible population—as opposed to only participants—and the members of the experimental group would have had to be permitted not to receive benefits (that is, the treatment would be only the offer of benefits, not their receipt). Unfortunately, this would permit an "experimental" estimate only of the effect of the intervention on the combined labor supply of participants and nonparticipants in the experimental group; an experimental estimate of the labor supply effect on only those who participated in the program would not be obtainable.[3] On the other hand, if randomization

2 Families were permitted to move above the break-even level and retain enrollment in the experiment, but they were not permitted to move off the program while below break-even. In the control group, on the other hand, many women below the AFDC break-even did not participate in AFDC, a characteristic common to most AFDC-eligible populations (see, for example, Moffitt, forthcoming)

3 If it could be safely assumed that the labor supply levels of nonparticipating experimentals were unaffected by the treatment, an estimate of the labor supply effect for participating experimentals alone could be obtained But this is an extra assumption that

were conducted only on recipients, the self-selection problems discussed previously would arise.

1967 Social Security Act Amendments. The 1967 amendments to the Social Security Act contained a number of provisions; one of the most prominent was the lowering of the nominal benefit-reduction rate in the AFDC program from 100 percent to 67 percent, a provision intended to provide work incentives similar to those of a negative income tax. Evaluation methodology was at a very early stage of development in the late 1960s and early 1970s, and hence only a few evaluations were conducted of the amendments, without the benefits of later methodological developments.

Appel (1972) and Smith (1974) examined aggregate data on the employment rates of welfare recipients in one state, Michigan, before and after the amendments. The object was to determine whether the employment of recipients was increased by the lowering of the benefit-reduction rate. Unfortunately, such a methodology is seriously flawed because the caseload after the amendments may have included new entrants who came onto the rolls to take advantage of the ability to work while on AFDC. Those new entrants may, for example, have been working the same amount while off AFDC and may have come onto the rolls simply to collect benefits.[4] Clearly the use of aggregate data does not permit the elimination of such compositional effects. A later study by Bell and Bushe (1975) used individual microdata, but only from a series of independent cross sections of the AFDC caseload conducted during the late 1960s and early 1970s. Again, the comparison of work-effort levels of recipients before and after the amendments in the successive waves of the microdata was contaminated by changes in composition and entry onto the rolls by workers. These studies show that it is panel data, not microdata per se, that are necessary to estimate entry effects in this type of analysis.

OBRA Evaluations. The 1981 OBRA legislation had a number of important features, but the major change in the AFDC benefit formula was the reinstitution of a nominal 100 percent benefit-reduction rate on

does not follow from the randomization itself, and must be justified independently of the randomization Indeed, in some evaluations the assumption is obviously unwarranted

4 This type of movement has been extensively discussed in the context of a negative income tax, for a lowering of the benefit-reduction rate is now known to create work disincentives through its attraction of new entrants at the same time that it creates work incentives for those initially on the rolls. See Levy (1979) and Moffitt (forthcoming).

earnings (at least after four months of work). An important question was whether this change would induce reductions in work effort among AFDC recipients, many of whom were working at the time of the legislation and could retain eligibility for benefits only if they stopped working.

Of the several evaluations of the legislation that were conducted, perhaps the best was that carried out by the Research Triangle Institute (RTI, 1983). The RTI design was based on a before-and-after examination of two cohorts. The first cohort was a nationally representative sample of AFDC recipients who were on the rolls just prior to the implementation of the OBRA legislation. The second (comparison) cohort was a similarly representative sample of AFDC recipients on the rolls one year prior to the first cohort. Both cohorts were followed over time, and their recipiency and work status were examined one year later to determine what percentage of recipients had left the rolls permanently and what percentage were still on the rolls but had moved from work to nonwork. The estimates of the OBRA effect on both the exit rate and the move-to-nonwork rate were based on a comparison of those rates in the "OBRA" cohort to those in the "pre-OBRA" cohort.

Although there are many issues surrounding this nonexperimental design, for present purposes the most important is its failure to permit the estimation of entry-rate effects of the legislation. Such effects should be expected to be present, for just as the 1967 amendments may have drawn new entrants onto the rolls who wished to work, the OBRA legislation should have been expected to discourage potential recipients who wished to work from joining the rolls. Just as the RTI design permitted the comparison of the exit rate in the OBRA cohort to a presumably "normal" exit rate in a pre-OBRA cohort, there should have been a comparison of the entry rate into AFDC in the OBRA period with a "normal" entry rate measured in the pre-OBRA period. That comparison could have been partially accomplished by the enrollment of replenishment samples each month to both the OBRA and pre-OBRA cohorts, as discussed later in this chapter. In any case, the RTI design thus did not permit the estimation of entry-rate effects nor, therefore, the net effects of the legislation on the AFDC caseload.[5]

MDRC Work-Welfare Experiments. The MDRC work-welfare experiments are reviewed in some detail by Greenberg and Wiseman in Chap-

5. For more detailed reviews of the OBRA evaluations, see Hutchens (1986) and Moffitt (1984, 1985).

ter 1 of this volume (see the references therein). The experiments were designed to evaluate the effects of welfare employment programs in eight different sites around the country. The programs were authorized by the 1981 OBRA legislation and provided various forms of job search, training, or work experience to AFDC and AFDC-UP recipients. The major goal of the evaluations was to estimate the effect of the various programs on the earnings of recipients, but effects on the exit rate from AFDC and the reentry rate onto the rolls ("recidivism") were also examined. The evaluation design was based on a randomization of individuals into experimental and control status, with experimentals receiving the new program and controls generally receiving existing services. However, the point of randomization varied across sites. In some, the randomization was conducted at the point of application for AFDC; in others, it was conducted at the point of certification of eligibility for AFDC, after some applicants had voluntarily dropped out of consideration or had been denied eligibility. In yet others, it was conducted at the point of reevaluation of eligibility, at the point of transition to mandatory program eligibility status, or after intensive recruitment for the program (see Greenberg and Wiseman, Chapter 1 of this volume.)[6]

Although the MDRC work-welfare experiments were admirably designed in many respects, they were little better than the prior evaluations discussed thus far in their provision for the estimation of entry effects. The randomization—at whichever point—was conducted only on AFDC applicants or recipients, and therefore no direct estimate of entry effects could be obtained. The estimates of effects on recidivism provided some information on reentry among those initially on the rolls, but it obviously did not provide information on entry-rate effects for those not on the rolls at the time of the experiment. Moreover, it is possible that exit-rate and recidivism effects in a national program would be different from those estimated in the experiment. If, for example, a national implementation of one of the programs tested were to increase the entry

6 In addition to creating some difficulties in interpreting the results across experiments, these differences in the points of randomization raise difficulties similar to those that would arise in income maintenance experiments that enroll eligibles rather than participants. For example, estimates of earnings effects for the subsample of an experimental group that actually receives and completes training are not possible without conducting an essentially nonexperimental analysis of differences in earnings within the experimental group. Note as well that the zero-effect assumption discussed in note 3 is unlikely to hold here These issues are also discussed by Greenberg and Wiseman.

rate, and if the new entrants had systematically different exit and recidivism rates than did those initially on the rolls, the long-run effects on such rates would be different from those estimated in the experiment. Likewise, as discussed previously for other evaluations, the earnings effects estimated in the MDRC experiments might be quite different from those that would obtain in a national program if the entry rate were to change and if those who were to enter (or not enter) a national program had systematically different earnings responses to training than did the recipients on the rolls during the experiments.

It is worth noting that the entry, exit, and net caseload effects of welfare employment programs of the types tested in the MDRC experiments are not predictable a priori and can only be determined by direct evaluation. The net effect is likely to be determined by the interplay of two separate forces.[7] On the one hand, the effect of such programs on the AFDC entry rate would depend on the net present value of the training or other employment opportunity to the recipient, taking into account the changes in current and future benefits, earnings, and work effort associated with the program. If that net present value were positive, the AFDC entry rate would tend to increase and the exit rate would tend to decrease; if it were negative, the opposite would tend to occur.[8] On the other hand, the positive earnings impact of the training itself would work to decrease future entry and increase exit, thus working in the opposite direction. That the net effect of these forces cannot be known without analysis serves to underscore the importance of estimating entry effects as a part of any complete evaluation of welfare employment programs.

Family Support Act of 1988. Evaluations of the Family Support Act have not been completed at this writing. However, the major components of the Act have a clear potential for entry-rate effects. The JOBS component of the Act, which mandates the implementation of certain types of welfare employment programs in all state AFDC programs, should affect entry rates and the caseload in a manner similar in type, though obviously not necessarily in magnitude, to those just discussed for welfare employment programs in general. As before, an estimate of these entry effects is required in order to obtain an estimate of the effect of the

7. I have discussed these issues in more detail previously (Moffitt, forthcoming).

8 Greenberg and Wiseman (Chapter 1 in this volume) report several negative net present values in the MDRC experiments.

JOBS program on the long-run AFDC caseload. Two other components of the legislation that may induce entry effects are those mandating transitional child-care and Medicaid coverage for up to 12 months after leaving the AFDC rolls. As discussed elsewhere (Moffitt and Wolfe, 1990), Medicaid extensions have the potential to induce entry onto the rolls because they make the program more attractive to potential recipients. Child-care extensions have the same possible effects. The magnitude of these entry effects would certainly be affected by the magnitude of the benefits made available in the extensions, the fraction of recipients receiving them, and, as a timing issue, how quickly knowledge of these extensions would percolate through low-income communities. Thus estimation of entry effects should permit an assessment of the time pattern of response as well as its short-run magnitude.

Evaluation Methods for Program Entry Effects

General Considerations

A formal mathematical model of entry onto and exit from the rolls of a welfare program is outlined in Moffitt (1990). That model is not necessary to an understanding of the discussion that follows but does provide a mathematical framework for it.

There are three significant implications of the model for the evaluation of program entry effects of an intervention. First, the decision-making process on the part of an eligible individual is made on the basis of the same factors that affect the process for decisions regarding program exit—namely, relative incomes, benefits, and other factors that are different on welfare and off welfare. This obvious conclusion is nevertheless important because it provides a prima facie case that program effects on exit and on entry will be correlated in sign, if not in magnitude. To hypothesize, for example, that an intervention such as a welfare employment program may affect exit but not entry requires hypothesizing that individuals do not follow the same decision-making process when deciding whether to go onto welfare as they do when deciding whether to go off. Or, to take another example, if Medicaid and child-care extensions are important enough to affect AFDC recipients' relative valuations of staying on versus going off AFDC, it should be expected that they will also be important enough to affect potential recipients' relative valuations of going on versus staying off. If such extensions are so minor compared to the other factors determining relative valuations that they do not significantly affect entry decisions, they are unlikely to be sufficiently important to affect exit decisions as well.

Second, the evaluation of the effects of interventions on exit rates—at least those interventions that affect the individual only while on welfare—can be conducted by examining either the effects of the offer of an employment program while on the rolls, or the effects of actual receipt of services from such a program; but the evaluation of the effects of such interventions on entry rates can be conducted only by examining the effects of an offer of an employment program should the individual choose to go on welfare.[9] This has important implications for the types of evaluations that must be conducted to estimate entry effects.

Third, if entry effects are present, the short-run and long-run effects of an intervention are almost certain to be different. An intervention that causes individuals either to enter the rolls in greater numbers or to decline to apply for welfare in greater numbers will affect the types of individuals who end up on the welfare rolls. This, in turn, may affect future exit rates and earnings impacts. Moreover, it should be expected that entry rates will change over time as the intervention remains in place, not only because knowledge of it will become more widespread but also because the program will be in place for potential recipients at earlier points in their lifetimes. When an intervention is first introduced, both existing recipients and nonrecipients will have already made many decisions on the basis of the program during their lifetimes prior to the intervention. Existing nonrecipients, for example, may have already made a considerable commitment to working while off welfare for several years and therefore may not find it advantageous at a late stage to consider entering AFDC even if it offers an attractive training opportunity. But women just reaching maturity, or just entering their first spell as a head of household, with little work experience or AFDC history, may find such an opportunity more attractive. When those women reach the same age as the older nonrecipients were at the time of the intervention, they may have higher stocks of human capital and higher wages. In the long run, obviously, all women will have matured with the intervention in place over their entire lifetimes.[10] Thus the immediate effects of an intervention may be quite different from the long-run effects.

9. Of course, the employment program may be mandatory if an individual is on welfare, but the welfare participation decision itself will always be voluntary.

10. Perhaps more obviously, existing recipients at the time of the intervention who have been on welfare for many years may have suffered considerable declines in their

Evaluation with Experimental Methods

Evaluations using experimental methods have a powerful attraction for program evaluation in general. Experiments have the twin advantages of providing a treatment variable that is assured to be at least conditionally independent of the unobservables in the sample, and of permitting the investigator to manipulate the treatment so as to be able to examine program variation that may not have naturally occurred in ongoing programs. Despite these advantages, however, experimental designs are less advantageously placed for the estimation of program entry effects than for the estimation of other outcomes. As noted earlier, experimental evaluation of program entry requires randomization of the offer, rather than the receipt, of program services within a sample of the eligible population—since the individual has to be on welfare to receive services—and it is unlikely that this will be feasible in a conventional experimental design where randomization takes place within sites but across individuals. Because the spread of knowledge within community-wide information networks is likely to be an important intervening variable in the program entry mechanism, individualized offers of treatments are unlikely to replicate adequately the program environment that would obtain in a permanent community-wide program.

Randomization across sites is therefore the only mechanism by which experimental methods are likely to generate adequate estimates of program entry effects—and, therefore, adequate estimates of net participation and caseload effects.[11] The difficulties with such saturation experiments are all of a practical rather than a theoretical nature. In saturation experiments the nature of the treatment is difficult to control and to standardize across sites; costs are likely to be very high; cooperation from local agencies may differ across sites in a way that generates selection bias; and, most important, it is generally quite difficult to obtain the sample sizes necessary for adequate statistical power of the treatment impact estimates. These and other difficulties with saturation-site designs have often been discussed in the literature (for example, by Orr,

labor force skills, but those women who are recipients some time after the intervention will have already had exposure to training programs and therefore may have higher skill levels.

11. Of course, the offer may be quite different to different groups of individuals within a site. For example, potential recipients with children greater than or less than 3 years of age will face different "offers" from the JOBS program But differences in the entry-rate responses between such groups cannot be separated from the effects of children themselves.

1988; see also Chapter 7 in this volume by Garfinkel, Manski, and Michalopoulos). The sample-size problem is particularly important, and it is made worse if the experimental design calls for stratification by type of area, as it often does. Unfortunately, the number of relatively large urban areas in the United States is finite and small.[12]

The problem of sample size is often addressed by sampling a relatively small number of areas but selecting pairs of sites matched on a set of observable site characteristics. Unfortunately, although matching procedures in general and paired-site matching procedures specifically have advantages in improving statistical efficiency and the precision of the estimates obtained in the evaluation, the sample sizes are still typically quite small and far less than required for adequate statistical power. It would be preferable, for example, to match many sites on each set of observable characteristics, not just two sites per set, but this is rarely feasible.

An additional difficulty with experimental estimation of program entry effects is the problem of limited experimental duration, which is discussed extensively in the literature on the income-maintenance and other large-scale social experiments in the 1970s. Long-run program entry effects are likely to be estimable only if the experiment is allowed to operate for a considerable period of time, thereby permitting knowledge of the change in the program to percolate through the eligible population and allowing the types of individuals who are on and off the rolls to change, as noted previously. Most experiments are not in place for sufficiently long periods for this to occur, even if the duration of the experiment were made part of the treatment design. Indeed, this is a case where estimated program entry effects are more likely to be accurate in the evaluation of an ongoing, permanent program than in the evaluation of one that is new and temporary.

12. For example, assuming equal numbers of treatment and control sites, the sample size of each necessary for a given power of the estimate of a difference in means is $2(\sigma t/\alpha)^2$, where σ is the standard deviation of the outcome variable, t is the level of t-statistic (power) desired, and α is the size of treatment impact to be detected. To take the case of earnings, if the desired t-statistic is 2, the expected treatment impact is \$1,000 annually—about the maximum obtained in the MDRC work/welfare experiments—and σ is, for example, \$3,000, the necessary sample size is 72, or 144 for the treatment and control sites combined. If σ were much lower, say \$1,000, a combined sample of 16 sites would be necessary. This exceeds the number of sites in virtually all past saturation experiments that have been implemented.

Evaluation with Nonexperimental Methods

Evaluation of entry effects with nonexperimental methods will in most cases require that there be natural variation in program services ("natural experiments").[13] Since the evaluation of entry effects requires variation in the offer, rather than the receipt, of services, this implies once again that cross-site or geographic differences are likely to be the major sources of such natural variation. It is well known that the major difficulty with such variation is that it may not be independent of other factors determining the outcome—in this case, entry rates. The valid estimation of treatment effects in nonexperimental analyses therefore requires, at minimum, "more" data than does estimation in experimental analyses because these other factors must be measured (see Moffitt, 1990, for a mathematical formulation). Natural variation is more likely to be dependent on variations in welfare histories of the populations, welfare benefit levels and other welfare characteristics, labor market characteristics, and the socioeconomic characteristics of the populations across the sites. At best, suitably exogenous variation in treatment characteristics across sites will be available only conditional on these types of variables, and therefore they must be measured and controlled in the analysis for valid impact estimates to be obtained. Such data collection generates an extra cost of nonexperimental methods relative to experimental methods, although one that is unlikely to outweigh the cost of operating experiments in as large a number of sites as could be analyzed nonexperimentally.[14] However, it remains the case that in many nonexperimental analyses the requisite data on welfare histories of the populations in different sites, for example, have not been collected, a sign that there must be significant costs involved. In addition, at worst, even significant data collection and control for site variables may not be sufficient to absorb all the unobservables on which cross-site treatment variation is dependent.[15]

13 Longitudinal data can in some circumstances also serve this function See Heckman and Robb (1985a, 1985b), Heckman and Hotz (1989), and Moffitt (1991)

14. An important determinant of the costs of operating saturation experiments is the extent to which the program costs themselves are borne by the experimental funders or the sites In the case of ongoing programs, the sites themselves are surely to pay for much of the cost, although their compliance in the experiment may require heavy subsidy See Chapter 2 by Hotz in this volume for a discussion of the JTPA experience in this regard

15. The relative data demands of experimental and nonexperimental analyses have been discussed previously in Moffitt (1991) That paper also summarizes the arguments

An additional cost of estimating entry effects, whether experimentally or nonexperimentally, is the cost of collection of data on nonrecipients themselves. Obviously the direct calculation of entry rates requires samples from that group. Unfortunately, household surveys in a large number of sites, particularly surveys that collect the detailed histories necessary for adequate nonexperimental control, are difficult and costly. This is in contrast to data collection for recipients, for whom administrative data will provide much, though not all, of the desired information. To be sure, large surveys are not necessarily required, for the sampling rate of the nonrecipient population need not be as high as that of the recipient population.[16] In addition, in some cases it may be possible to utilize information on nonrecipients from publicly available data sets such as the Current Population Survey (CPS), the Survey of Income and Program Participation (SIPP), and similar household surveys, possibly matched to administrative records from the Social Security and AFDC systems. However, in many cases the sample sizes in such data sets will be too small for reliable estimation, particularly if the population under study is highly restricted. An additional disadvantage of these data sets is that they do not contain as detailed information on participation histories as could be obtained with a new household survey, although an advantage is that they are available historically—new household surveys that attempt to capture characteristics of past states of nonrecipiency are unlikely to be reliable. In addition, sole use of such publicly available data sets for program evaluation invariably leads to delay in the analysis, a disadvantage from the policymakers' point of view.

Analysis with Program Data Only. In light of the potential expense and difficulty of collecting information on the nonrecipient population, it is worth considering the possibility of evaluating entry effects with administrative data from the welfare system alone. Many evaluations rely heavily if not exclusively on such administrative data, a prominent example being the RTI OBRA evaluation discussed previously. Some evaluations also supplement the administrative data with household survey informa-

of Heckman and Hotz (1989) regarding the methods that can be used to ascertain when "enough" variables are available to warrant confidence in nonexperimental treatment-effect estimates.

16. The literature on choice-based sampling has demonstrated that different sampling rates of such populations can even improve statistical efficiency and can, in any case, be adjusted for in the analysis of the data. See Manski and Lerman (1977), Manski and McFadden (1981), Cosslett (1981), and, more recently, Imbens (1990).

tion on the recipients while they are still on the rolls or after they have left them. In any event, none of the major evaluations using administrative data have attempted the estimation of entry effects, and therefore it is not clear that such data can be used for that purpose.

Fortunately, administrative data can be used to address the estimation of entry effects by making use of data on new entrants to the welfare rolls. Subsequent to an intervention that varies across sites, variations in the numbers of new entrants across those sites will reflect, in part, any entry-rate effects that may have been induced by the intervention. However, variations in the numbers of new entrants may also reflect variations in the size of the eligible population. Nevertheless, treating that size as an unobservable leads directly to a consideration of whether the variation in treatment is independent of that unobservable, possibly after conditioning on a set of observables that are correlated with that size. In the lucky circumstance that treatment variation is independent of that size, variations in the numbers of new entrants across sites will permit valid estimates of entry effects. It is more probable, of course, that conditioning on additional variables will be required. One source of such variables will usually be aggregate data on population and other site characteristics available from published or unpublished governmental data bases. But another obvious source of such variables will be past values of the size of the welfare caseload and of new entrants into that caseload. A simple control for prior size of caseload is sufficient to standardize roughly for cross-site variation in numbers of new entrants by size of the population, and additional historical series on caseload sizes can be used to control for variations in growth rates of the caseload. Time-series data on the numbers of new entrants, either alone or as a percentage of the caseload, should provide additional leverage in controlling for the unobservables determining post-treatment variation in numbers of new entrants. In the final analysis the issue will be whether controlling for such observables leaves the conditional treatment variation independent of the remaining unobservables affecting post-intervention entry.[17]

The use of historical administrative information on caseload and entry sizes involves the same type of analysis as that generally conducted in studies of "caseload modeling." Caseload modeling has been con-

17. This procedure is exactly analogous to the type of procedure that must be followed with individual data in any nonexperimental analysis.

ducted for many years (see Lyon et al., 1975, for an early example) and has been used for a few recent program evaluations—see Chapter 1 in this volume by Greenberg and Wiseman, the studies discussed in Chapter 3 by Fishman and Weinberg, or the recent evaluation of the Massachusetts ET program by O'Neill (1990). However, it is the exception rather than the rule in these evaluations that entry, rather than caseload size per se, is the focus of the analysis.

The experience with caseload modeling has not been very favorable to date, for estimated caseload models are often found to be unreliable, nonrobust, and sensitive to specification. It is not clear whether this instability is a result of estimation with data from only one city or one state, as most caseload models have been, or a result of the inherent loss of information involved in the use of aggregate data.

The latter difficulty can be surmounted by replacing the traditional method of caseload modeling with an analysis of individual microdata drawn from the administrative records of the welfare system. Samples of records drawn from the records of all the sites in the analysis, and drawn for several periods in the past and into the future, would furnish information on individual characteristics that has been lacking in traditional caseload models. Moreover, collection of individual welfare histories from administrative records would permit an individual-specific analysis of entry (and exit) that is not possible with aggregate data.[18]

To date, the only evaluation effort to use a similar approach to this is the RTI OBRA evaluation, which used administrative information on the welfare experiences of two cohorts of recipients, one pre-OBRA and one post-OBRA, both of whose experiences were followed over a one-year period through administrative records. As noted previously, the evaluation failed to examine entry effects, but this could be remedied by collecting data on new entrants, or what are sometimes called "replenishment" samples. In addition, a complete analysis of this type would require two other significant extensions of the OBRA evaluation frame. First, more extensive histories on the welfare recipients should be collected, because one-year histories are unlikely to furnish sufficient control to guarantee conditional exogeneity of the typical treatment variables that are available. Second, more than two cohorts will in general be required to control for time-varying effects. Indeed, ideally a periodic sample of new entrants in every month or every few months

18 Use of microdata does not mean that the aggregate data must be ignored, the aggregate information can still be used in conjunction with the microdata See Lancaster and Imbens (1989) for an econometric discussion of this issue

for several periods prior to and subsequent to the intervention should be collected, with each cohort of new entrants then followed through the administrative records for the duration of the study.[19] One of the additional products that could be constructed from such a data base would be a sample of "ever on" individuals, namely, a sample of those members of a particular eligible population who were ever on welfare in the past. This sample could implicitly serve as a proxy for the unobservable discussed at the beginning of this section, namely, the population of eligibles from which new entrants are drawn.

Obtaining Adequate Cross-Sectional Treatment Variation. It is worth noting in conclusion to this discussion of nonexperimental methods for estimating entry effects that it is particularly important for all nonexperimental evaluations, both those on entry and those on exit, and both those with representative-population samples and those with administrative data only, for the data collection effort to include a detailed characterization of the treatment offered in each site. Such a characterization is generally obtained in fair detail in experimental studies because the analysts are more directly involved in the administration of the treatment, but a similar level of detailed characterization is generally missing from nonexperimental studies. The RTI evaluation, for example, was significantly hampered by the lack of any cross-sectional treatment variation whatsoever, even though the implementation of the OBRA rules did vary across the states; the RTI evaluation therefore relied solely on a time-series, before-and-after method of inference. Other evaluations (including randomized experiments) sometimes include a process analysis, but its results are rarely as integrated into the impact analysis as is necessary. This issue will be particularly important in any nonexperimental evaluation efforts conducted for the effects of the Family Support Act, for that act permits considerable cross-sectional variation in both the types of programs offered under the JOBS legislation and the character of the Medicaid and child-care extensions offered in each area.

Entry Effects and the Evaluation of Earnings Impacts

As noted earlier in the review of past evaluations, the neglect of entry effects in past evaluation efforts could have a deleterious effect on the earnings impacts estimated in those studies as well. The conditions un-

19 Supplementing these data with administrative information on earnings would add significant power to the analysis as well.

der which a problem will arise are demonstrated in formal terms in Moffitt (1990), but the intuition for the results given there is not difficult to obtain.

There are three conditions that must hold for a problem to arise. The first is that the effect of the treatment on earnings must vary across different individuals. This is a rather plausible eventuality, for different individuals no doubt differ in their past skill levels and therefore in their ability to take advantage of the training and other opportunities offered by welfare employment programs, for example. Indeed, discussions of "creaming," "targeting," and differential impact of employment programs for the more disadvantaged and the less disadvantaged implicitly presume that treatment impacts do vary across individuals.

The second condition is that an individual's likelihood of being on welfare is correlated with his or her ability to gain from the treatment. It is also quite plausible that this might be the case, for if the individual's ability to gain from the program is correlated with her underlying employability, for example, it will certainly be related to her probability of being on AFDC in the first place because that probability will be highly related to employability.

The third condition is that the evaluation is conducted by examining only those on welfare at the time of the intervention. As noted in the review of past evaluations, this has been the case for most experimental and nonexperimental evaluations.

If these three conditions hold, a problem will arise in the estimation of earnings effects because the estimated magnitude of those effects will change if the participation rate does, and that participation rate will change if there are entry effects in response to the intervention. The magnitude of estimated earnings effects will be dependent on the magnitude of the participation rate because individuals with a higher probability of participating in AFDC have different treatment responses, as discussed in the second condition above. Thus, for example, if an intervention involving a successful employment program draws new individuals onto the rolls and if, for illustration, those new entrants have higher labor force skills and are more employable than those on the rolls at the time of the intervention, then the average earnings impact of the program among those on the rolls after the intervention will presumably gradually rise. If those newly entering the rolls have lower levels of skills than those initially on, the opposite would result.[20]

20 Here it is implicitly assumed that the earnings impact of the employment program is positively correlated with the level of skills and employability. This is not obvious, the

A simple means of addressing this problem is to build into the design of the evaluation a means by which the dependence of mean estimated treatment responses in different sites on the participation rate can be explicitly estimated. Either by stratifying the sample on the initial participation rate or by examining participation-rate variations that naturally occur in the sample as part of the analysis, an evaluation with sufficient numbers of sites would permit the correlation of site-specific earnings effects and participation rates. Then, once an analysis of entry has been conducted and estimates of the effect of the intervention on participation rates have been obtained, the change in the mean earnings effect can also be estimated.

Conclusions

The neglect of program entry considerations is one of the most striking characteristics of the majority of program evaluations in the area of welfare and training. Yet, as stressed in this chapter, a careful consideration of program entry effects is required to properly estimate long-run effects on the caseload, on the costs of the program, and, possibly, on earnings and labor supply as well. The evaluation of entry effects, as well as the more general evaluation of long-run effects on program participation, requires somewhat different methods than does the evaluation of effects on program exit or on earnings conditional upon program participation. Although experimental methods are still feasible, they require randomization across sites rather than across individuals and are subject to limited-duration problems. Consequently, the relative advantage of experimental methods is considerably reduced for the estimation of entry and participation-rate effects. Nonexperimental analyses, even those which use administrative data only, are more likely to be feasible for such estimation.

Why the study of entry effects has been so extensively neglected in the program evaluation literature is not completely clear. It is possible that the experimental paradigm, which serves as a conceptual model for most evaluations including those conducted with nonexperimental methods, has focused attention away from the study of entry rate ef-

reverse may be the case, for example If so, the direction of change in mean treatment effect magnitudes will be reversed. The point is the same, indeed, this serves to demonstrate the importance of estimating these effects to determine the direction of the bias fects. Because conventional experimental methods are not particularly well suited for the analysis of entry, they do not lead analysts ordinarily to consider entry effects at the time of evaluation design.

It is also worth stressing that experimental methods may of course still have a comparative advantage in addressing other outcomes such as exit and earnings impacts, although there is considerable disagreement in the literature even on this issue. But presuming that there are cases where the powerful advantages of experimental methods can be exploited, there is still a strong case for the supplementation of those methods with nonexperimental analyses of entry. It is quite possible that a combined experimental-nonexperimental analysis could provide treatment-effect estimates that complement each other, and could produce net impact estimates that are considerably stronger than the sum of the parts that could be obtained from either type of evaluation conducted individually.

References

Appel, Gary Louis. 1972. "Effects of a Financial Incentive on AFDC Employment: Michigan's Experience between July 1969 and 1970." Minneapolis: Institute for Interdisciplinary Studies. Photocopy.

Bell, Winifred, and Dennis M. Bushe. 1975. *Neglecting the Many, Helping the Few: The Impact of the 1967 AFDC Work Incentives.* New York: Center for Studies in Income Maintenance Policy.

Bjorklund, Anders. 1988. "What Experiments Are Needed for Manpower Policy?" *Journal of Human Resources,* 23 (Spring): 267–277.

Bjorklund, Anders, and Robert Moffitt. 1987. "Estimation of Wage Gains and Welfare Gains in Self-Selection Models." *Review of Economics and Statistics* 69 (February): 42–49.

Cosslett, Stephen. 1981. "Efficient Estimation of Discrete Choice Models." In *Structural Analysis of Discrete Choice Models,* ed. Charles F. Manski and Daniel McFadden. Cambridge, Mass.: MIT Press.

Goldberger, Arthur. 1972a. "Selection Bias in Evaluating Treatment Effects: Some Formal Illustrations." Discussion Paper 123-72, Institute for Research on Poverty, University of Wisconsin–Madison.

——— 1972b. "Selection Bias in Evaluating Treatment Effects: The Case of Interaction." Discussion Paper 129-72, Institute for Research on Poverty, University of Wisconsin–Madison.

Heckman, James, and V. Joseph Hotz. 1989 "Choosing among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training." *Journal of the American Statistical Association,* 84 (December): 862–874.

Heckman, James J., and Richard Robb. 1985a. "Alternative Methods for Evaluating the Impact of Interventions: An Overview." *Journal of Econometrics,* 30 (October/November): 239–267.

——— 1985b. "Alternative Methods for Evaluating the Impact of Interventions." In *Longitudinal Analysis of Labor Market Data,* ed. James Heckman and Burton Singer. Cambridge: Cambridge University Press.

Hutchens, Robert. 1986. "The Effects of the Omnibus Budget Reconciliation Act of 1981 on AFDC Recipients: A Review of Studies." In *Research in Labor Economics,* ed. Ronald Ehrenberg. Vol. 8, Pt. B. Greenwich, Conn.: JAI Press.

Imbens, Guido. 1990. "An Efficient Method of Moments Estimator for Discrete Choice Models with Choice-Based Sampling." Discussion Paper 9009, Department of Economics, Tilburg University, The Netherlands.

Lancaster, Tony, and Guido Imbens. 1989. "Uses of Marginal Information in Econometrics." Manuscript, Department of Economics, Brown University. Photocopy.

Levy, Frank. 1979. "The Labor Supply of Female Heads, or, AFDC Work Incentives Don't Work Too Well." *Journal of Human Resources,* 14 (Winter): 76–97.

Lyon, David W., C. Peter Rydell, and Mark Menchik. 1975. "Welfare Policy Research for New York City: Findings on the Dynamics of Dependency." Paper P-5565, Rand Corporation, Santa Monica, Calif.

Manski, Charles F., and Steven Lerman. 1977. "The Estimation of Choice Probabilities from Choice-Based Samples." *Econometrica,* 45 (November): 1977–1988.

Manski, Charles, and Daniel McFadden. 1981. "Alternative Estimators and Sample Designs for Discrete Choice Analysis." In *Structural Analysis of Discrete Data with Econometric Applications,* ed. Manski and McFadden. Cambridge, Mass.: MIT Press.

Moffitt, Robert. 1984. "Assessing the Effects of the 1981 Federal AFDC Legislation on the Work Effort of Women Heading Households: A Framework for Analysis and the Evidence to Date." Discussion Paper 742A-84, Institute for Research on Poverty, University of Wisconsin–Madison.

——— 1985. "Evaluating the Effects of Changes in AFDC: Methodological Issues and Challenges." *Journal of Policy Analysis and Management,* 4 (Summer): 537–553.

——— 1989. "Comment on Heckman and Hotz." *Journal of the American Statistical Association,* 84 (December): 877–878.

——— 1990. "Issues in the Estimation of the Impact of an Intervention on Welfare Participation and Earnings." Mimeographed, Brown University.

——— 1991. "Program Evaluation with Nonexperimental Data." *Evaluation Review,* 15 (June): 291–314.

——— Forthcoming. "Incentive Effects of the U.S. Welfare System: A Review." *Journal of Economic Literature.*

Moffitt, Robert, and Kenneth Kehrer. 1981. "The Effect of Tax and Transfer Programs on Labor Supply: The Evidence from the Income Maintenance Experiments." In *Research in Labor Economics.* ed. Ronald Ehrenberg. Greenwich, Conn.: JAI Press.

Moffitt, Robert, and Barbara Wolfe. 1990. "The Effects of Medicaid on AFDC Participation and Labor Supply." Special Report 49, Institute for Research on Poverty, University of Wisconsin–Madison.

O'Neill, June. 1990. *Work and Welfare in Massachusetts: An Evaluation of the ET Program.* Boston: Pioneer Institute for Public Policy Research. Photocopy.

Orr, Larry. 1988. "Choosing between Macroexperiments and Microexperiments." In *Social Experimentation,* ed. Jerry Hausman and David Wise. Chicago: University of Chicago Press, for the National Bureau of Economic Research.

Rangarajan, Anuradha. 1989. "The Effect of the AFDC Program on Human Capital Accumulation." Ph.D. diss., Department of Economics, Brown University.

Research Triangle Institute. 1983. "Final Report: Evaluation of the 1981 AFDC Amendments." Research Triangle Institute, Research Triangle Park, North Carolina. Photocopy.

Smith, Vernon. 1974. "Welfare Work Incentives: The Earnings Exemption and Its Impact upon AFDC Employment, Earnings, and Program Cost." Michigan Department of Social Services, Lansing. Photocopy.

SRI International. 1983. *Final Report of the Seattle-Denver Income Maintenance Experiment:* Vol. 1. *Design and Results.* Menlo Park, Calif.: SRI International.