1995

Can Economic Development Programs Be Evaluated?

Timothy J. Bartik
W.E. Upjohn Institute, bartik@upjohn.org

Richard D. Bingham
Cleveland State University

Upjohn Institute Working Paper No. 95-29

****Published Version****

In Dilemmas of Urban Economic Development: Issues in Theory and Practice, Richard D. Bingham, Robert Mier, editors. Thousand Oaks, Calif.: Sage, 1997, pp. 246-290

Can Economic Development Programs be Evaluated?

Upjohn Institute Staff Working Paper 95-29

Timothy J. Bartik
W. E. Upjohn Institute
for Employment Research

and

Richard D. Bingham
Cleveland State University

Pepared for: Richard D. Bingham and Robert Mier, *Significant Issues in Urban Economic Development* (Newbury Park, CA: Sage).

Can Economic Development Programs be Evaluated?

Abstract

The question addressed in this paper seems simple: Can economic development programs be evaluated? But the answer is not simple because of the nature of evaluation. To determine a program's effectiveness requires a sophisticated evaluation because it requires the evaluator to distinguish changes due to the program from changes due to nonprogram factors. The evaluator must focus on the outcomes caused by the program rather than the program's procedures.

Evaluations can be divided into two categories—process or formative evaluations and outcome, impact, or summative evaluations. Process evaluations focus on how a program is delivered. Impact evaluations focus on the program's results. Although process evaluations are important, the focus of this chapter is on program outcomes—thus the concern with impact evaluations; however, both types of evaluations need to be defined.

I. Types of Evaluations

Practitioners often disagree with professional evaluators, particularly academic evaluators, about what is a satisfactory evaluation. For some practitioners a response to a question asked of a recipient of a state or local economic development program concerning the number of jobs "saved" or "created" by government assistance constitutes "proof" that the program is worthwhile. For the evaluator, however, such responses may be suspect.

Some of the disagreements over evaluation methodologies between practitioners and academics stem from conflicting opinions about what constitutes evaluation. One of the major text books in the field defines evaluation as,

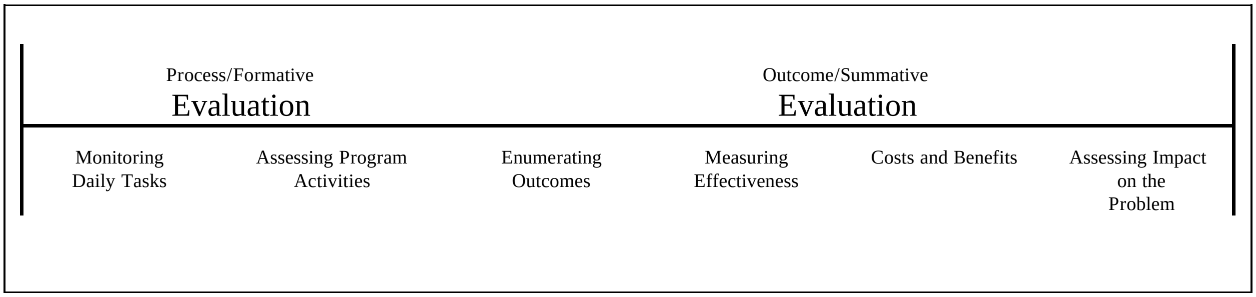
> *the systematic application of social research procedures in assessing the conceptualization and design, implementation, and utility of social intervention programs*. (Rossi and Freeman, 1985: p. 19)

The book then elaborates:

> *evaluation research involves the use of social research methodologies to judge and to improve the planning, monitoring, effectiveness and efficiency [emphasis ours] of health, education, welfare, and other human services programs*. (p. 19)

A fairly broad definition as to what constitutes evaluation is reasonable. Thus the concepts of planning, monitoring, effectiveness, and efficiency have a place in the evaluation of economic development programs.

It is helpful to look at evaluation as a continuum moving from the simplest form of evaluation, monitoring daily tasks, to the more complex, assessing impact on the problem. Such a continuum is illustrated below:



The six points illustrated on the continuum represent different levels of evaluation with each level building on the previous one. These evaluation activities take place in roughly the same sequence as the implementation of the program. First, tasks are monitored, then activities are assessed, outcomes are enumerated, the effectiveness of programs are measured, and finally, a judgement is made as to whether the problem has been reduced (Trisko and League, 1978). The lower level functions are best at providing information about how a program can be improved. The higher level evaluation techniques determine if a program works (e.g., actually creates jobs). This partially explains the economic developer's proclivity for process evaluations (discussed later).

The lowest level of evaluation, monitoring daily tasks, simply examines the internal working of a program. The monitoring activity examines questions about the management of the organization such as: Are contractual obligations being met? Are staff members working where and when they should? Is the program administratively sound? Are daily tasks carried out efficiently? Are staff adequately trained for their jobs? Monitoring activities are designed to uncover and deal with management problems. Work analysis, resource expenditure studies, management audits, procedural overhauls, and financial audits are considered monitoring-level evaluation activities.

The next level of evaluation is assessing program activities. Here the characteristics of current activities are identified. In assessing program activities the following types of questions are asked: What activities are taking place? Who is the target of activity (businesses, cities, etc.), and with what problems or needs? How well is the program implemented? Could it be done more efficiently? Are clients satisfied? Does the program have a favorable image? Thus, this level of evaluation assesses numbers and types of programs services, the recipients of those services, program efficiency, and other similar characteristics of the program.

The preceding levels of evaluation are process or formative evaluations. The next level, enumerating outcomes, is the first of the outcome or summative evaluations. Enumerating outcomes allows one to determine whether program objectives—the immediate short-term outcomes—have been achieved. Typical questions at this level might be: What is the result of the activities described in the process evaluation? What happened to the target population? How is it different from before? Have unanticipated outcomes occurred and are they desirable? Have program objectives been achieved? To ascertain what the results of the program are, some over-time measures are needed. How are the program recipients different from the way they were before?

Merely enumerating outcomes, however, does not allow one to attribute changes that occurred simultaneously with the program's existence to the impact of the programs. This is essentially the problem solved by measuring program effectiveness. Effectiveness measurement tells one whether program goals have been accomplished. The questions of causality are specifically addressed: What would have happened in the absence of the program? Does the program work? What are the other factors that may have contributed to changes in the recipients? To answer these questions a cause and effect relationship must be established between the program and the outcome. Did the tax abatement "cause" an increase in employment in the target company?

However, simply because a program is shown to be 3substantatively effective -- that is, it actually does create jobs -- does not mean that the program should have ever been implemented. Cost-benefit analysis allows one to determine if the program benefits outweigh the program costs. Often they do not. For example, it is questionable whether the public costs of subsidies for the construction of several large auto plants will ever be recovered (Jones and Bachelor, 1984; Elder and Lind, 1987). Thus cost-benefit analysis simply asks: Do costs of the program outweigh the benefits of the program?

Finally, because a program has been shown to be both substantively effective and cost-effective does not mean that there is an improvement in the problem situation. The reality is that most programs do not have the necessary resources to have a measurable impact on the problem. Assessing the impact requires answers to the following types of questions: What changes are evident in the problem? Has the problem been reduced as a result of the program? What new knowledge has been generated for society about the problem or the ways to solve it? From a practical standpoint it is difficult to answer these questions even if the program had adequate resources. This level of evaluation moves from the program to policy—and true policy analysis is enormously expensive to carry out. Also, so many external variables exist that it is virtually impossible to determine real impact on social problems resulting from policy intervention.

These six levels of evaluation, although they have no strict boundaries, provide a framework for assessing the quality of evaluations that have been conducted of economic development programs. As was shown earlier, in order to effectively "prove" that a program accomplishes its goals, the evaluation must be at the highest two levels of evaluation—measuring effectiveness or, assessing impact. Simply enumerating outcomes is not sufficient.

II. The Problem of Outcome Evaluation

The question: Can economic development programs be evaluated? is hard because it is difficult (but necessary) to determine what would have happened to the program participants if the program did not exist. The evaluator wants to compare what actually happened with what "would have happened if the world had been exactly the same as it was except that the program had not been implemented" (Hatry, Winnie, and Fisk, 1981: p. 25). But because it is impossible to determine exactly what "would have happened if…" the answer is to use procedures that give an approximate idea of what would have happened.

A. Threats to Validity

One reason it is hard to determine what "would have happened if…" is that changes taking place in the world may affect the program participants so as to make it look as though the program works when it actually does not, or, conversely, makes it look as though the program does not work when it actually does. For example, historical events occurring during the program may cause changes in program participants that could appear to be the result of the program. In the economic development arena, one common historical event that can lead to such misinterpretations is the business cycle. Suppose a city gives a firm a property tax abatement which coincides with the bottoming out of a national recession. During the following three years the firm completes its physical expansion, does well in the market and increases its employment by 50 percent. The difficulty for the evaluator is determining how much of the employment increase is due to the abatement, how much is due to the national economic recovery, and how much is due to the firm's performance.

Maturation of a firm or an industry can cause a similar problem for the evaluator. The growth pattern of a new firm or industry approximates an "s" shaped curve. During the initial years the firm grows slowly as it innovates and perfects its product. This period is followed by a time of rapid growth when the firm's product holds a competitive advantage in the market and the firm grows rapidly. Then, as other companies enter the market, the firm's market share (and employment) level off and may even begin to decline. The evaluation problem is that if a firm uses economic development assistance just before beginning its "natural" period of rapid growth, it will be difficult to distinguish that natural growth in employment from the growth resulting from the economic development assistance.

Another reason it is hard to determine what "would have happened if…" is that program participants may differ from non-participants because of how the program selects participants, or how participants select themselves. Almost all economic development programs are selective in that only a minority of firms participate, and the firms that participate are not chosen randomly. The firms that self-select themselves for participation, or are selected by program managers, might have performed differently from non-participants without the program. Evaluations of the program will be biased if they attribute these performance differences to the program.

In theory this "selection bias" in evaluations of economic development programs could either make the programs look better than they are, or worse than they are. In practice, for most economic development programs, this "selection bias" will make the programs look better than they are. Firms applying for economic development assistance will frequently be firms that are aggressively seeking to expand. The desire of these firms to expand is why they applied for a property tax abatement, a training grant, help with modernization, or help with exporting. Even without economic development assistance, such aggressive firms would be more likely to expand, train, modernize, and export.

For some economic development programs, selection bias may make the programs look worse than they are. For example, enterprise zone programs select for assistance all firms located in neighborhoods that are hostile to business, with high crime, poor infrastructure, inadequate labor skills, and low consumer demand. Firms in such neighborhoods would be likely to do much more poorly than firms in average neighborhoods without the program. Even a successful enterprise zone program will find it difficult to stimulate firms in these depressed neighborhoods to out-perform firms in other neighborhoods.

In the lexicon of program evaluation, the problems caused by history, maturation, or selection are referred to as "threats to validity." Evaluators typically utilize different evaluation designs to attempt to overcome the threats to validity. Three common designs are illustrated in Figure 1.

AT THIS TIME FIGURE 1 IS AVAILABLE IN HARD COPY ONLY.

B. Models to Assess Outcomes

The design shown in Figure 1 (A), the one-group pretest-posttest design, is one of the most commonly used evaluation designs. It is not a powerful design and is subject to most of the traditional validity problems. Here the group receiving the program is measured before the initiation of the program (A_1) and again after the program is completed (A_2). The difference scores are then examined and any improvement (A_2 - A_1) is usually attributed to the impact of the program. The major drawback of this design is that changes in the participating firms may be caused by other events and not the program. The longer the time lapse between the preprogram and postprogram measurements, the more likely it is that other variables beside the program affected the postprogram measurement.

The design in Figure 1 (B), the pretest-posttest comparison group design, is a substantial improvement over the simple one-group design because the evaluator attempts to create a comparison group that is as close as possible to the program recipients. Both groups are measured before and after the program and their differences are compared $[(A_2-A_1) - (B_2-B_1)]$.

(Note: the percentage change might be used rather than the absolute amount of change.) The use of a comparison group can alleviate many of the threats to validity as both groups are subject to the same external events. The validity of the design depends on how closely the comparison group resembles the program recipient group.

The pretest-posttest control group design (Figure 1 (C)) is the most powerful and "truly scientific" evaluation design presented here. The only significant difference between this design and the comparison group design is that firms are assigned to the program and control groups randomly. The firms in the program group receive program assistance and those in the control group do not. The key is random assignment. If the number of subjects is sufficiently large, random assignment implies that the characteristics of subjects in both groups are likely to be quite similar prior to the initiation of the program. This initial similarity, and the fact that both groups will experience the same historical events, mature at the same rate, etc., reasonably assures that any difference between the two groups on the post-program measure will be the result of the program.[1]

C. Other Approaches to Measuring Outcomes

These different formal evaluation designs all use "objective" evidence from the program itself to infer the extent to which the program "causes" various outcomes. For economic development programs, there are other, more informal approaches to determine the program's effects. One approach is to survey firms that are program participants, and ask them whether the program caused various changes in the firm's actions. This "subjective" approach seems unscientific, but that may be because natural scientists seldom are able to ask their experimental subjects (e.g. rats or electrons) about the causation issue.

One problem with surveying firms about causation is that firms may have difficulty giving precise and accurate responses. Even if the program surveys the relevant decision-maker in the firm, that decision-maker may find it difficult to give a precise quantitative assessment of the effect of the program on the probability of making an investment decision. A second problem with surveying firms is that some firms may have incentives to exaggerate the program's effects on the firm's behavior. The economic development program may only provide assistance if a firm says the assistance is essential to the decision, in which case admitting that the assistance was not essential might raise legal issues. The firm's managers may also believe that future assistance is more likely if they give a favorable view of the program.

Despite these problems, surveying firms about causation may yield useful evidence in cases where firms do not have strong incentives to lie. If firms at least try to tell the truth, the

---

[1]Although a number of authors in the evaluation field use the terms *control group* and *comparison group* interchangeably, they are not. Control groups are formed by the process of randomization. Comparison groups are groups that are matched to be comparable in important respects to the experimental group. In this chapter the distinction between control groups and comparison groups is strictly maintained.

responses to surveys about causation may provide a rough guide to the effectiveness of the economic development program. Comparing responses from several programs with similar target groups of firms and similar goals may allow some assessment of which program is more effective.

Another approach to estimating the effects of an economic development program is to simulate plausible firm reactions to the program's change in the incentives facing the firm. This simulation of the change in firm behavior would be based on some explicit or implicit model of how firms behave. Such a model would presumably be based on some other source of data. For example, suppose an economic development program lowers wage costs per hour by a wage subsidy, or lowers effective wages by increasing labor productivity. The resulting average effect on firm employment might be inferred from previous studies of how firms' labor demand decisions respond to changes in wages. The accuracy of such simulations depends on the extent to which the simulation model captures the main relevant features of the economic development program. In the present example, the simulation of the employment effects of a wage subsidy would be less accurate if firms also view the wage subsidy as symbolizing a friendly local "business climate", an aspect of the program that is difficult to model.

The preceding discussion has implicitly assumed that the evaluator of the economic development program already knows the outcomes to be measured. But determining the outcomes to be measured is a difficult art. It is important to distinguish between "ultimate outcomes," the outcomes one would really like to measure, and "proximate outcomes," the outcomes that actually can be measured. For example, suppose, in evaluating an export assistance program, that the program's effect on a community's total employment is viewed as the "ultimate outcome." But this might be too hard to measure given the small size of the program, so an evaluation might want to look at the program's effects on jobs in the assisted businesses. If this is too hard to measure, the evaluation might look at the program's effects on exports in the assisted businesses, and if that is too hard to measure, it might examine the program's effects on a firm's efforts to export.

Another important distinction is between outcomes at the local level and outcomes at some broader geographic scale, such as the national level. A program that subsidizes new business investment in Mississippi is likely to have quite different effects from Mississippi's perspective than from a national perspective, because much of any increased investment in Mississippi will come at the expense of reduced business investment in other states. Which geographic perspective to appropriate depends on who will be using the evaluation and for what purposes.

An even more complicated issue is determining what should be the ultimate outcomes by which one might judge economic development programs, and how to value these ultimate outcomes. Suppose new jobs is viewed as an appropriate ultimate outcome. Then an immediate question is how to place a value on the program's job creation, to allow a comparison with the costs of the program. The value of this job creation is likely to depend on who gets the jobs from the program, and what they would have been doing with their time if they had not obtained the jobs created by the program. Thus, the process of trying to determine the real value of a program's ultimate outcomes will usually lead to some redefinition of these outcomes.

Most economists and other professional public policy analysts would argue that economic development programs should ideally be subjected to cost-benefit analysis:

> *Cost-benefit analysis is an intuitively easy process to conceptualize and appreciate. One gathers all of the costs of providing a good or service and weighs those costs against the dollar value of all the subsequent benefits provided by the goods or service. If the benefits outweigh the costs, the good or service should be continued; if the costs of providing the service exceed the benefits obtained, the service should be terminated.* (Bingham and Felbinger, 1989: p. 207)

Thus with cost-benefit analysis both the costs and benefits of a program are expressed in dollar terms. For example, in the case of a tax abatement, the cost of the program would include the value of the taxes abated. Benefits would include taxes generated for the locality by the salaries of the new employees, multiplier effect of salaries paid to employees living in the jurisdiction, and such.

Although cost-benefit analysis is conceptually the correct way to approach economic development evaluation or any other public policy evaluation, it can be difficult to apply in practice. For example, it is difficult to know exactly what dollar value is to be attached to a job that goes to a person who otherwise would be committing crimes versus a job to a person who otherwise would be "unemployed" but taking care of their children at home. But even if such questions are difficult to answer conclusively, asking such questions at least directs the evaluator's attention to important issues. In the present example, asking the question of the value of a job directs the evaluator's attention to the important issue of determining who gets the jobs from a particular economic development program.

III. A Critical Review of Some Examples of Different Types of Economic Development Evaluations

It is impossible to provide a comprehensive review of all economic development evaluations in this chapter, so this section will critique examples of different types of economic development evaluations. Types of economic development evaluations considered here include: process evaluations; before and after evaluations; survey evaluations; evaluations based on models of firm behavior; comparison group evaluations; randomized control group evaluations; and evaluations that use models of community impact.

Process Evaluations

Many state economic development programs have been the subject of process evaluations by state audit agencies or other such "good government" groups. These process evaluations typically focus on dull but important issues such as the need for better planning, increased targeting of resources, and improved monitoring of program activities.

A good example of a high quality process evaluation is the recent evaluation of Virginia's economic development programs by the Virginia Joint Legislative Audit and Review Commission (Joint Legislative Audit and Review Commission, 1991). Among other things, this report recommended that Virginia should try to reduce duplication of services between small business development centers and technology transfer programs. The report recommends greater targeting of programs: industrial training assistance should be targeted more towards higher wage industries, retention calls targeted towards firms in key local industries that the local economy is having trouble retaining, and marketing efforts targeted towards areas of the U.S. and the world from which Virginia has successfully recruited companies in the past. The report recommends improved program monitoring: better monitoring of the activities of marketing staff, better efforts to measure the payback of industrial training programs, and more standardized measures of the activities of small business assistance programs.

These recommendations seem sensible. But the report does not provide much evidence on whether Virginia's economic development programs make a difference to the assisted firms or to Virginia's economy.

Enumerating Outcomes -- Before and After Evaluations

Another common evaluation approach is to enumerate outcomes and to assume that any new activities of assisted firms must be attributable to the economic development program. For example, a recent study by the Pennsylvania Economy League, an independent government "watchdog" group, of the Northeast Tier Ben Franklin Technology Center, concluded that "through December 1990, a total of 3817 jobs have been created and a total of 4420 jobs have been retained as a direct result of [the Northeast Tier's] Ben Franklin Center administered projects" (Pennsylvania Economy League, May 1992). These job numbers are derived by assuming that any job created or retained in a project with Northeast Tier assistance was due to the program.

More sophisticated outcome enumeration studies at some point admit some doubts about causation. Knowledge Systems and Research, Inc., in a 1991 study of New York State's Entrepreneurial Assistance Program, claims at one point that

> *an estimate of business creation impact is provided by the survey of clients 2-3 years after initial assistance was received by EAP where the reported percentage of new businesses among start-up clients is 30.8%…Other measures of*

> *entrepreneurial assistance impacts for FY 1989 are: business expansions -- 68 (24.3%); businesses saved -- 60 (21.5%); jobs created (other than business operators) -- 36*. (Knowledge Systems and Research, Inc., 1991, pp. 4-7, 4-8)

Later on in the report, however, the consultant admits that "it cannot be determined in the scope of this evaluation how many of these jobs have been or will be created but for EAP assistance." (p. 4-11) In other words, what were previously called "impacts" of the program may not really be "impacts."

Enumerating Outcomes -- Survey Evaluations

Many evaluations of economic development programs have asked clients of the programs to assess their impact. The reliability and usefulness of this survey evidence varies a great deal from one study to another, depending on the type of economic development program and how the survey results are used. Mt. Auburn Associates and the Corporation for Enterprise Development, in a 1990 study of New York State's business financing programs, surveyed 186 firms that had received loans either from the state Job Development Authority, the Urban Development Corporation, or the Science and Technology Foundation. The report used surveys to examine the effects of the financing in several ways:

> *For 41% of the surveyed firms, availability of state financing was the deciding factor in making the investment in the community. Without state financing, these firms would have canceled the investment, moved to another community, or gone out of business. For another 40% of the surveyed firms, state financing affected the scale or timing of the investment decision. Eighteen percent of the surveyed firms indicated they would not have altered their investment decision had the state financing not been available.*(p. 17)…
>
> *We asked the companies surveyed to provide their employment level as of November 30, 1989, and to answer hypothetically what their employment* ***would have been on the same date in the absence of the financing****. [emphasis in original] The difference between actual and hypothetical employment levels counts those jobs retained and created* ***attributable to the financing****. [emphasis in original]… Since the question related to employment in the absence of the financing is hypothetical, the answers are not precise. Yet, only the firm's owner of management fully understands the factors responsible for employment growth or decline. The company itself is the only source of information to address the issue of "attribution"…*
>
> *Based on this job impact measure, the level of job retention and creation was 6,876 jobs…The difference between this figure and the job growth of 7,137…results from firms reporting that some jobs created could not be attributed to the financing. The job impact measure recognizes the job retention due to*

financing and specifies if the financing was responsible for job creation. (pp.23-25). (Mt. Auburn Associates, 1990).

Although this survey approach is preferable to giving the program credit for everything that happens at firms that receive financing, there may be some incentives for firms to misrepresent the effects of the financing. For many financing programs, laws or program rules require that loans go to firms only in cases where the loan is necessary for the project to proceed. If a firm said that a project would have proceeded without the state financing, and the firm's response became public, serious political or legal problems could arise.

Another study that used survey evidence to assess program effectiveness is the previously mentioned study of New York State's Entrepreneurial Assistance Program. EAP actually consisted of two separate programs, support centers and development centers, with support centers apparently providing more formal training of potential entrepreneurs. In response to the question "To what degree did the center help you achieve your goal?" a majority of the clients of both types of centers felt the center helped at least somewhat but "support center clients were most positive, with development center clients less convinced that the program was helpful." (Knowledge Systems and Research, Inc., 1991, p. 4-9)

This relative ranking of these two types of centers is convincing. Center clients were not required to claim that the service was essential in order to receive service. If the entrepreneurial assistance was not helpful, what incentive is there for clients to lie and claim that the service was helpful? If the service was not helpful, clients would not perceive any reason to maneuver to keep the programs alive. In contrast, for programs that provide cash assistance, there is always an incentive for firms to keep the programs alive and available to the firm, because cash assistance is always helpful to the firm even if the assistance does not change the firm's behavior.

Even if EAP clients give truthful survey responses, it is unclear how a client saying that EAP helped in achieving the client's goals translates into a precise quantitative effect on the probability that the client will start a small business or expand a small business. But because two programs with similar goals and clientele are considered, the relative ratings of the two programs may give a rough qualitative idea of the relative effectiveness of the two programs. The comparison suggests that formal training sessions are a useful component of an entrepreneurial training program.

A third example of an evaluation using survey evidence is a large scale 1992 study, conducted by Mt. Auburn Associates, Brian Bosworth, Brandon Roberts, and Arthur D. Little, of Ohio's Edison Technology Centers program (Mt. Auburn et al, December 1992). This study included both surveys of firms involved with the Edison Centers, and focus groups with some of these firms. Overall, 40% of Edison member firms indicated strong effects of the Edison programs on either process, skills, customers, cooperative relationships, or business practices. Perhaps as important, the reported impact varied a great deal by center:

> *On a firm by firm basis, EBTC [Edison BioTechnology Center] appears to be the Center with the most impact among its members. It had the highest percentage in several impact categories-- 60 percent said EBTC had a strong impact on at least one aspect of business operations, 33% said there were Ohio employment impacts, and 74% said that if EBTC did not exist present and/or future performance would suffer…There is very high enthusiasm among the member firms…*
>
> ** CAMP [Cleveland Advanced Manufacturing program] works intensively with far more Ohio-based firms than any other Center…A higher percentage of members said that CAMP aided in manufacturing process improvements and in business practice improvements than was so for any other Center…Enthusiasm among interviewees and focus group members for CAMP services is very high.*
>
> ** EABC [Edison Animal Biotechnology Center] has only worked with three Ohio firms over the course of its existence. Its impact has been minimal.*
>
> ** Measures of IAMS' [Institute of Advanced Manufacturing Sciences] current performance must take into account that the current incarnation of IAMS is less than two years old…A lower percentage of members said that IAMS had a strong impact on firm operations, or any positive impact, than for any other Center…Focus group members tended to have less enthusiasm for belonging to the Center than was so at other Centers."*
>
> *The analysis in the individual chapters and a review of the survey tables indicates that the economic effectiveness of the four other Centers tends to be between those discussed above.* (Mt. Auburn et al, 1992, pp. 12-6 to 12-7)

These comparative rankings of the different Edison Centers are convincing for the reasons previously stated—it is unclear why Center members would have an incentive to say that the services have an impact if they are useless, and the differences in survey results across centers are likely to correspond to a rough qualitative ranking. In addition, because the survey results are backed up by focus groups, they also seem more convincing. It requires a more extended effort to lie throughout a focus group discussion. Also, the focus group discussions provide some insight into why different centers are more or less effective.

Enumerating Outcomes -- Evaluations Based on Models of Firm Behavior

Another approach to enumerating outcomes is to use some model of firm behavior that simulates whether a program made some difference to a firm's behavior. One example of a study using such a model is a 1989 study by the Illinois Auditor General of the economic development programs of the Illinois Department of Commerce and Community Affairs (DCCA). This report blasted DCCA for poor management of its loan and grant programs to firms—DCCA was criticized for poor documentation of the programs, and an unclear decision-making process for making grants. These criticisms appear to be well-founded. But the report also analyzed the effectiveness of DCCA's loan and grants, based on a model of when a state subsidy will alter a firm's location or expansion decision, and this part of the report is more questionable.

The Auditor General's report made two main adjustments to decide when a DCCA loan or grant was responsible for a firm's location or expansion decision. First, the Auditor General made an adjustment based on the firm's rate of return before and after the DCCA loan or grant:

> *If a firm had a rate of return at the time of the subsidy that was either negative (the firm was losing money) or below industry averages, we concluded that the firm needed assistance. If subsequent to the subsidy the firm's rate of return became positive and comparable to the industry average, we concluded the subsidy had been effective. If these conditions were absent, we concluded the subsidy was unwarranted or ineffective or both.* (Illinois Auditor General, 1989, p. 98)

Based on an audit of a sample of 35 DCCA projects, the Auditor General concluded that in 12 projects DCCA's assistance was ineffective, making up 21% of DCCA's claimed jobs for all 35 projects.

Second, the Auditor General adjusted DCCA's job creation numbers based on the percentage of capital provided by DCCA.

> *In our calculations we prorated the number of jobs created on the basis of the proportion of capital provided. If DCCA contributed 10 percent of the capital, then 10 percent of the jobs created were attributed to DCCA's activity.* (p. 7)

Because on average DCCA supplied about 13% of the capital for its projects, this eliminated about 87% of DCCA's claimed job creation. These two adjustments, plus other adjustments made by the Auditor General, reduced DCCA's claimed creation of 7501 jobs in this sample of 35 projects to only 608 jobs.

The Auditor General's model of the effectiveness of economic development loans or grants is questionable. The two adjustments do roughly correspond to factors that influence the effectiveness of economic development subsidies, but the Auditor General's specific job creation numbers could easily be wildly inaccurate. With respect to the first adjustment, even if a firm is making above normal rates of return, a subsidy may increase profits in Illinois over profits in some alternative location, changing that location decision. An effect on location or expansion decisions is more likely if the subsidy has a larger effect on rates of return, but there is unlikely to be any fixed cutoff point determining when a subsidy will make a difference. With respect to the second adjustment, one would not think that the subsidy's effect on the probability of a location or expansion decision would be exactly proportional to the percent of capital supplied, although perhaps the more capital supplied, the more likely the subsidy is to affect the decision.

If the loan or guarantee alters the firms profits enough so that location or expansion in Illinois is preferred to the alternative of not locating or expanding in Illinois, whereas before it was not, then 100% of the jobs involved are attributable to the subsidy, regardless of the percentage of capital supplied.

Another program evaluation that uses an implicit model of business behavior is a recent "self-evaluation" done of Michigan's Capital Access Program (Rohde et al, 1990). Under the Capital Access Program (CAP), a business borrowing money and the bank each pay a fee of 1.5% to 3.5% of the loan, with the bank's fee typically passed on to the borrower. These fees, and a state contribution of 150% of the combined bank/borrower fees, are paid into a separate reserve fund for each bank participating in the program. This reserve fund is available to cover losses in loans made under the program; a bank's losses beyond its own reserve fund are the bank's responsibility. CAP is designed to encourage banks to take greater risks with small business loans, but not too much risk. Most banks aim for less than a 1% loss rate on small business loans. Under CAP, loss rates of 10% or so can easily be tolerated, and still be profitable to the bank.

CAP is designed so that in a competitive banking market, rational business borrowers and rational banks would not want to participate in CAP unless the loan is above normal risk, because of the extra costs associated with CAP loans. Rohde and the other staff running CAP conclude in their report that the above normal loss rate of CAP shows that the program is working as intended:

> *There is strong evidence from a variety of sources that the Capital Access Program is causing banks to make loans that they otherwise would not make.*
>
> *The most compelling evidence comes from the…loss rate under the program. The loss rate under the program has been running probably at least 7 times…a normal bank loss rate. If banks, on average, when using the program take 7 or more times the risk they normally take without the program, it is a reasonable conclusion that the program is making a difference in causing them to take this added risk.* (Rohde et al, 1990, p. 46)

This is strong evidence, but it really comes from a model of how banks and borrowers behave in a competitive banking market rather than any direct empirical proof that loans are being made under the program that otherwise would not have been made.

Some of Bartik's (1991, 1992) work on economic development has essentially argued that "a cost is a cost," and financial subsidies to firms should have similar effects regardless of whether provided through regular taxes, or through special tax subsidies. For example, Bartik (1992) argued that the annual cost per job created through reductions in general state and local business taxes is somewhere between $2000 and $11,000 per year. In the absence of evidence to the contrary, there is some plausibility that special tax and other financial subsidies to firms will have similar average effects. For example, if a branch plant with 1000 workers is given a financial subsidy equivalent to $1 million annually—$1000 per worker per year—it could be argued that the subsidy is unlikely on average to affect the probability of the plant choosing that location by more than 50%. General business tax reductions seem to cost at least $2000 per year per job created, so it seems unlikely that on average a special tax subsidy for a typical firm would be much more effective than that. Of course, in any particular case, the subsidy either tips the decision or it doesn't, but we are unlikely ever to know whether the subsidy made a difference for certain.

There are two problems with this line of reasoning. First, economic development financial subsidies may have more than a strictly economic value to the firm—they may symbolize a good business climate. Second, some of the evidence on effects of different types of changes in costs seems contradictory. For example, the average estimated effect in most business location studies of a 10% reduction in wages is not much greater than the effect of a 10% reduction in state and local business taxes, even though wages are a much greater proportion of costs than state and local business taxes. Bartik (1991, pp. 49-52) suggests some reasons why estimated effects of wages on business location decisions may be biased towards zero, but this pattern still raises some doubts about the "model" that all cost changes have similar effects on business location.

Evaluations Using a Comparison Group, Without Random Selection

Another way to measure the effects of economic development programs is to compare the performance of program participants with a comparison group of non-participants. By definition, this comparison group will not be randomly chosen, but there will be some effort to make the group comparable to the program participants.

One example of a non-random comparison evaluation is the study by Price Waterhouse of the U.S. Small Business Administration's (SBA's) 7(a) Guarantee Business Loan Program (Price Waterhouse, March 18, 1992). This study compared recipients of SBA guaranteed loans in FY 1985 with a comparison group of firms drawn from SBA's Master Establishment List, which is derived from Dun and Bradstreet records and yellow page listings. In selecting the comparison group, Price Waterhouse tried to match recipients to non-recipients by two-digit industry, number of employees, and 10 SBA regions. Price Waterhouse did a phone survey of recipients and the comparison group in 1991, asking about 1984 and 1989 revenues and employment. The study found that recipients were more likely to be in business in 1989 than non-recipients, and had more growth from 1984-89 in revenues and employment. But SBA loan guarantee recipients were also more likely than comparison group firms to be start-up firms, and many firms in the comparison group said they didn't obtain an SBA loan because they didn't need one. These facts imply that perhaps SBA loan guarantee recipients would have been more aggressive in expanding even if this SBA program had never existed. Price Waterhouse is careful to note that we cannot tell what would have happened to SBA loan guarantee recipients without the loan.

Another study with a comparison group is a study by Public Policy Associates and Brandon Roberts of Oregon's small business service programs (Public Policy Associates and Brandon Roberts, October 14, 1992). This study included surveys of assisted small businesses and a general survey of Oregon small businesses. The report notes that users of state small business services did better on increases in sales, profitability and jobs during the 1989-92 period. This finding is not emphasized in the report. This deemphasis seems appropriate, because again it may be that users of services may have been more aggressive about expanding even without the services.

A somewhat different example of a comparison group evaluation is Leslie Papke's study of the effects of Indiana's enterprise zone program. Her study differs from most other studies of economic development program because her study seeks to look directly at overall performance of some geographic area (in this case, enterprise zone neighborhoods) compared to other geographic areas. Ideally all economic development programs would be evaluated based on direct evidence on how they affect their target geographic area, whether that is a state, metropolitan area, city, or neighborhood. But such "community-wide" impact evaluations are not possible for most economic development programs because the programs are so small relative to their geographic area of interest that any estimated impact is likely to be spurious. Enterprise zone programs at least claim to be large enough relative to their chosen neighborhoods to have a detectable overall community economic impact.

Papke finds that after zone designation, zone areas experience a reduction in unemployment claims compared to non-zone areas. One problem with her study is that she never describes in any detail how she selects the comparison group of non-zone areas in Indiana. Hence, it is unclear whether the non-zone areas are truly comparable to the zone areas.

Another innovative comparison group evaluation of economic development programs was done by Holzer et. al. on Michigan's former program of industrial training grants for manufacturing firms undergoing modernization. The evaluation compares firms that received training grants in 1988 and 1989 with firms that applied for such grants, but too late in the state's fiscal year to receive a grant. Apparently the state's modernization training grant program generally funded all eligible firms until they ran out of that year's budget. Holzer and his colleagues surveyed both groups of firms in 1990. They found that firms that received grants did more training afterwards than non-grantees, and their product scrappage rates were reduced more.

The Holzer evaluation could be criticized using the argument that firms that apply too late for state grants are less capable than firms that apply on time. But the two groups of firms did not seem to differ significantly in observed characteristics. In addition, one could argue that firms have some incentive to claim that the grants led to reductions in scrappage rates. But the incentive to lie is relatively low because firms that received training grants were ineligible for future grants.

Evaluations Using Randomly-Assigned Control Group

Another evaluation approach is to compare the performance of program participants to a randomly assigned control group of firms. The random assignment ensures that the control group will be similar to program participants on all relevant characteristics, both observed and unobserved.

The only economic development evaluation that has used random assignment is an Abt study of two experiments, sponsored by DOL, of entrepreneurship training for unemployment insurance (UI) recipients in Washington state and Massachusetts. The experiments first ran orientation sessions explaining the entrepreneurship training program; the 2 to 4% of UI recipients who expressed interest in such training were then randomly assigned to a treatment group and a control group. Both experiments showed significant increases in self-employment in the treatment group compared to the control group, with no significant increase in the probability of business failure. In Massachusetts, 47% of the treatment group ended up with some self-employment experience, compared to 29% of the controls, whereas in Washington state, 52% of the treatment group entered self-employment, compared to 27% of the controls. As this experiment shows, many individuals who received no entrepreneurial assistance still started up their own business. This emphasizes the point that one cannot attribute all business start-ups associated with an economic development program to the program itself (and clearly illustrates the threat to validity posed by self selection).

Evaluation Models of Community Impact

Even if an evaluation study has determined the effects of an economic development program on firms participating in the program, an important remaining issue is what impact the change in behavior of the program participants has on the overall local community. As mentioned above, because most economic development programs are so small relative to the community, any impact will be difficult to detect, but the impact may be large relative to the size of the program.

To analyze how changes in participant firms in turn affect the community, an economic impact model must trace out how the initial employment and other impacts on participant firms lead to multiplier effects on the activities of other firms, as well as effects on unemployment and labor force participation, and wage rates. A fiscal impact model that relates these economic impacts to effects on local taxes and public service costs is also needed.

A wide variety of models already seek to do economic impact and fiscal impact analyses. One such model that has been specifically adapted to economic development programs is the University of Illinois-Chicago's adaptation of George Treyz's well-known REMI economic impact model. Wiewel, Persky and other UI-Chicago researchers have adapted the REMI model to trace out how a given economic development project will have fiscal impacts on Chicago and economic impacts on Chicago residents (Persky, Felsenstein and Wiewel, 1993; Wiewel, Persky, and Felsenstein, 1994).

One important adjustment in community impact models is that the community effects of assistance to export-base firms—firms that export their product outside the regional economy—are likely to be greater than the community effects of assistance to locally-oriented firms. Economic development programs that help export-base firms are unlikely to have negative effects on the sales of other local firms. In contrast, economic development programs that help some firms that mainly sell to the local market are likely to have some negative effects on other firms that also sell to the local market. Helping locally-oriented firms may have some positive effects on the local economy, because some of the expansion in these firms' sales may come at the expense of imports to the local economy, not other local firms. But these positive effects on the local economy will probably be less than the positive effects of assistance to export-base firms.

Overall Assessment of Current Evaluations

This review of economic development evaluations may be misleading because examples of several types of evaluations were provided, which could give the impression that each type of evaluation has been done with roughly similar frequency. The reality is that the vast majority of existing evaluations are process evaluations. A process evaluation is the standard type of evaluation done by a state audit agency or a good government group. There also are a growing number of economic development evaluations that use qualitative ratings from surveying the program's client firms. Only a few evaluations of economic development programs use comparison groups. In most of these evaluations, the chosen comparison group does not seem comparable. Only the Holzer study of Michigan's economic development training program uses a comparison group taken from unsuccessful program applicants, which at least controls for the more aggressive firms being more likely to apply. Only the Abt study of entrepreneurial training programs uses a randomly assigned control group.

The emphasis on process evaluations and surveys enumerating outcomes is unfortunate because evaluations with a good comparison group or a control group are likely to do the best job of measuring the true quantitative effects of economic development programs. Surveys of client firms provide some qualitative evidence on these programs' effectiveness. This qualitative evidence may be particularly valuable when comparing similar programs. But surveys of clients alone are not a reliable means of quantifying exactly what difference the economic development program has made to a firm's performance. Such surveys cannot really address the crucial issue of whether the quantitative effects of economic development programs are worth their costs.

This is an era that demands that government programs show that their benefits are greater than their costs. An important but puzzling question is why, in this atmosphere of accountability, there have not been more well-done comparison or control group studies of economic development programs? That issue is considered in the next section.

IV. Why Aren't There A Greater Number of More Sophisticated Economic Development Evaluations?

If only evaluations with comparable groups can really tell us whether economic development programs are successful, why aren't there more such evaluations? There are six reasons for such a situation. All of these reasons amount to saying that whatever the social benefits and costs of more sophisticated evaluations, such evaluations are not perceived as offering sufficient benefits to justify their costs from the viewpoint of those groups that would usually have to authorize, pay for, or conduct such evaluations.

The first reason for the lack of sophisticated evaluations is that such evaluations are difficult to do, compared either to no evaluation at all, or compared to evaluations focused on process or on subjective responses to survey questions. Evaluations with a comparable group by definition require careful procedures to select this comparison group. Such evaluations also require collection of extensive quantitative data over a period of time from both the firms participating in the economic development evaluation, and the comparison group. In order to allow for a comparable group, and allow for the collection of baseline data, the evaluation should ideally be built into the program design.

These data collection and design efforts may not only be expensive in direct budgetary costs, but may also require extensive administrative time and be disruptive to the staff setting up the program. It is much easier to do no evaluation at all, or to simply interview some program staff or program clients later on and ask them what they think of the program.

A reason that randomized control group evaluations are unpopular is that such evaluations require explicitly denying services to specific firms. Of course economic development services are implicitly denied to many firms by providing so little in funds to economic development programs. In addition, many experiments with medical treatments or welfare programs do use randomized control group assignment to explicitly deny services to some individuals. But somehow the explicit denial of services seems more bothersome when those being denied services are firms, perhaps because this violates our sense of how a free market system should work.

A third reason for the paucity of comparison/control group evaluations is that more rigorous evaluations will have a disproportionate part of their benefits go to groups other than those paying for the evaluation. Hard quantitative evidence on the effectiveness of a particular approach to economic development will have benefits to all state and local areas, not just to the state or local area that has the program and is funding the evaluation.

A fourth reason for the unpopularity of rigorous evaluations is that such evaluations too often seem to avoid telling program administrators how they can improve their program. Outcome evaluations are frequently written as if the program is a "black box," and shy away from trying to determine why a program was or was not successful. Knowing whether the program was or was not successful is of course the most important issue, but the program administrators want to know how to improve the program. A process evaluation would seem to offer some clues as to how to improve the program, even if the evaluation by itself does not document what the program really accomplished. Process and outcome evaluations should not be thought of as being mutually exclusive alternatives, but as being complementary.

A fifth reason for the lack of sophisticated evaluations is that many of the groups that typically do economic development evaluations, such as state audit agencies, and some consulting firms, do not seem to be oriented towards evaluation studies that use sophisticated methods. State audit agencies frequently do not have staff who are trained in how to do studies that correct for selection bias due to a non-randomly selected comparison group. Nor do state agencies typically have staff with much familiarity in how to set up an experiment with a randomly assigned control group.

A final, and perhaps most crucial reason that more rigorous evaluations are rare, is that program administrators fear the political consequences of a negative evaluation. If a program is not evaluated, one can always claim success. A process evaluation or a survey evaluation is subjective enough that it may be easier to manipulate the evaluation process or reinterpret the results to make the program look better. But if a study shows that firms participating in the economic development program, compared to a truly comparable group, show no difference in performance, then it is difficult to argue that the program works.

These fears of the possible negative political consequences of evaluation are realistic given the American political tradition of suspicion of government. Christopher Jencks, a well-known sociologist at Northwestern University, recently commented on the peculiar American attitude toward government programs:

> *In every other society with which I am familiar, the political system assumes that when a program is not working you try to improve it. Only in the U.S., where we doubt that government can ever do anything right, do we assume that if a program stumbles in its first couple of years we ought to terminate it. No wonder we have so few successful programs. It's like deciding that if babies get sick they should be thrown away.* (Jencks, 1992-1993, p. 12)

As a result, economic development program managers, or indeed managers of any government program in the U.S., are perhaps justified in following the advice of economists Gary Burtless and Robert Haveman: "If you advocate a particular policy reform or innovation, do not press to have it tested." (Burtless and Haveman, 1984)

The real issue in encouraging more sophisticated evaluations of economic development programs is how to change the perceptions of economic development program managers, state legislators, and others who would have to authorize such evaluations about the benefits and costs of more rigorous evaluations. Some thoughts about how to do so will be offered in the conclusion.

V. A Suggested Approach to Economic Development Evaluation

Three types of economic development evaluations are needed:

** Community evaluations—evaluations that directly estimate how some economic development initiative affects an overall community;

** Firm evaluations—evaluations that directly estimate how an economic development program affects individual firms;

** Community Impact evaluations—evaluations that calculate the impact of some assumed change in firm behavior on a community.

Community evaluations

As was mentioned, directly estimating the effect of some economic development program on a community is only feasible when the program is "large" relative to the size of the community. Enterprise zones are one case where such evaluations seem feasible. In addition, it seems feasible to make such an evaluation in the case of large scale public works investment, such as highways, in rural counties (for an example, see Rephann, 1992).

Community evaluations should ideally use random assignment. Such random assignment is feasible when only a small number of deserving communities can be helped with the program. For example, the Clinton Administration's recently-enacted enterprise zone law is providing assistance to 9 "empowerment zones" and 95 "enterprise communities." Although the selection of these "zones" and "communities" was not based on random assignment, it would have been possible to do so, for many more communities applied for designation than could be accommodated under the program. If random assignment were used, a simple comparison of changes in "treatment" and "control" communities would allow an unbiased assessment of the effects of the program.

If random assignment is not used, then community evaluations should compare assisted community with other eligible communities that applied for the program but were not assisted. Using other eligible, applicant communities increases the similarity between the "treatment" communities and "comparison" communities. Differences will still remain. To control for such differences, researchers will have to make a careful effort to identify other variables which might predict the economic performance of treatment and comparison communities. Past economic performance is one good control variables.

Firm Evaluations

For economic development programs that are too small to plausibly have a detectable direct effect on the community, evaluators should focus on the effects of the program on the performance of assisted firms. Most economic development programs fall in this category.

Such evaluations of effects on individual firms should ideally use random assignment to divide firms into treatment and control groups. Random assignment means that a simple comparison of the performance of the treatment group and control group firms should give an unbiased estimate of the effects of the program, as treatment and control group firms must on average be similar.

Random assignment is most feasible in cases where there is significant excess demand by many similar firms for a program, the program provides similar assistance to all assisted firms, and the timeliness of the assistance is not the key issue. This situation is likely to occur in cases where a program provides cash (loans or grants, etc.) or near-cash assistance (for example, dollars for training) to many small and medium sized firms—many small firms are likely to want the assistance even if they have to wait a bit, the assistance provided to firms is similar, and the program is likely to need some type of application deadline and decision procedure to determine who gets the assistance. There may also be excess demand by similar firms for "classroom" style educational and information programs, for example in entrepreneurial assistance training programs.

Excess demand by similar firms for similar services is much less likely in industrial extension service programs—these programs provide customized assistance to individual firms, and strive to provide the assistance as quickly as possible. Asking industrial extension customers to wait for random assignment is likely to hurt the reputation and effectiveness of the program. Excess demand by similar firms is also less likely for programs that provide cash or near-cash assistance to large firms. Each large firm client and location or expansion decision is likely to be unique. Because each large firm decision is likely to have sizable economic impacts, few government officials are going to want to risk a particular case being assigned to the control group.

Even if there is excess demand, random assignment may not be politically or administratively feasible. A second-best approach is to use, as a comparison group, rejected firms that applied for assistance under the program. This procedure at least controls for the greater aggressiveness and growth orientation of applicant firms compared to non-applicant firms. The analysis comparing assisted firms with rejected applicants should include variables controlling for other factors affecting the economic performance of firms, particularly if such factors were part of the process by which the program selected which applicants to fund.

This caveat is extremely important. Because of the enormous number of products manufactured, distributed, and sold in the United States, random assignment of establishments into experimental and control groups will almost never produce similar product mixes (unless the number of establishments is extremely large). This is important because industries differ widely in terms of their product and profit life cycles (see Markusen, 1985) and because of a given industry's place on the business cycle (leading versus lagging indicators).

It is also important because for many jurisdictions and programs the number of firms receiving assistance is quite small. Thus, even if it were possible to randomly assign firms into treatment and control groups, it is unlikely that the two groups would be similar.

In both of the above instances the only realistic alternative is to develop a matched-pair comparison group whereby each firm receiving the program is carefully matched with a similar firm not receiving the programs. Extreme care must be taken to insure that each comparison group firm is closely matched to its equivalent control group firm on all relevant variables.

Thus a matched-pair comparison group might be a more-than-satisfactory alternative. There are a number of ways to identify such like firms. They can be identified utilizing secondary sources such as by combining the use of ES202 data, industrial directories, and annual reports. The simplest and most effective way, however, is simply to ask the firms recieving the programs. Survey experience with these firms has shown that they can easily identify similar firms.

If the program either does not have excess demand by similar firms for similar services, or the evaluator is unable to get data on rejected applicants, then other firms may also be used as a comparison group. Because firms that did not apply to the program are likely to be less aggressive and growth-oriented than applicant firms, the researcher should try to include in the empirical analysis variables that control for a firm's growth orientation—for example, prior growth of the firm. The analysis will also be helped if the researcher can find variables that help predict whether a firm is assisted by the program, but are plausibly not otherwise strongly correlated with the firm's economic performance. For example, how close a firm is to the program's offices could plausibly affect the likelihood of a firm receiving assistance, but might not be strongly related to the firm's performance otherwise. If variables that predict program participation can be identified, but do not directly affect performance, then these variables can be used as "instrumental variables" to control for the non-random selection of firms for participation in the program. This is not the place to explain in detail how such instrumental variable selection bias correction procedures work—this topic is covered in many econometrics texts.

In cases where no comparison group data are available—the program is targeted at large firms, or no funds are available for collecting data on unassisted firms- surveys of assisted firms may be useful. Surveys collecting qualitative information from assisted firms may also be a useful supplement to more quantitative information available from control group surveys. Surveys asking assisted firms to assess the impact of the program are likely to be most believable if the surveys are anonymous, and if the program is designed so that firms do not have any legal or political pressure to claim that the program was effective. Qualitative surveys will be more useful if several similar programs can be compared using similar surveys. National standards for survey questions would be helpful. These national standards would allow state and local policymakers to compare their economic development program's performance with similar programs in other areas.

Community Impact Evaluations

If one can know or assume some effect of an economic development program on individual firms, it still is of interest to estimate or simulate what impact that change in individual firm behavior will have on the community's fiscal and economic situation. Even if this community impact is not large enough to be detectable, it could be large compared to the size of the economic development program.

A good economic and fiscal impact study should be sensitive to at least the following five issues:

1. Export-base versus non-export base companies. As was mentioned earlier, economic development assistance to non-export base companies, companies that primarily sell locally, is likely to have some adverse effects on the market share and performance of other locally-oriented firms. These adverse competitive effects on unassisted local firms are less likely for economic development assistance to export base firms.

2. Population in-migration. Increases in local employment caused by economic development programs are likely to lead to in-migration. Research reviewed by Bartik (1993, 1991) indicates that for every 5 jobs created in a metropolitan economy, 4 jobs are likely to go to in-migrants in the long-run (five years or so). This increase in population reduces the employment benefits of economic development programs to the original local residents, and is likely to raise the fiscal costs of economic development programs.

3. Adverse effects due to increased costs and other congestion effects. As an area grows, land and housing prices increase, forcing up wages. In addition, growth may cause greater traffic congestion and environmental problems. All these greater costs are likely to discourage employment growth in some firms. Hence, the initial increase in local employment growth due to some economic development program is likely to be offset to some extent in the long-run due to these higher costs.

4. Public service costs. Additional business activity and population will create the need for additional public expenditure. Too many analyses of economic development seem to pretend that growth only increases the tax base without increasing spending needs.

5. Public capital costs. Some of the most expensive public expenditures required by growth are likely to be the capital costs associated with building new infrastructure, or retrofitting

old infrastructure to accommodate growth. These increased costs for additional population and business will often be more expensive, on a per person or per employee basis, than the capital costs for the current population and employment base. Retrofitting infrastructure is often expensive. In addition, the existing infrastructure may have already been paid for by past generations, and may have been partially funded by federal grants.

VI. Conclusion

Can economic development programs be evaluated? Our answer is yes they can—and furthermore, some decent evaluations have already been done. Good quantitative evaluations of economic development programs require the use of a comparable group of unassisted communities or firms. Such quantitative studies can be usefully supplemented by well-designed qualitative surveys of assisted firms, done so that results can be compared across similar programs.

But although some reasonably good evaluations have been done, such evaluations are far too rare. Good evaluations are rare because the policymakers who would have to approve such evaluations perceive too much political risk and/or cost in economic development evaluations.

To encourage more good evaluations, the perceived benefits and costs of economic development evaluations need to be changed. This can be done in three ways:

(1) <u>Federal funding and standards</u>. Federal funding of evaluations, along with federal funding of demonstration programs, would be a useful carrot for encouraging state and local governments to do more evaluation. In addition, federally-funded economic development programs should have the "stick" of requiring rigorous evaluation.

To ensure that evaluations are rigorous and comparable across different state and local areas, the federal government should sponsor a cooperative effort with state and local economic development organizations to develop standards for high-quality economic development evaluations. Standards are particularly important for qualitative surveys, whose value is enhanced by having evidence from similar surveys of similar programs. The results of such rigorous evaluations should be disseminated with federal funds, through publications, conferences, and electronic networks.

Which federal agency should be charged with promoting economic development evaluations is difficult to decide. One logical agency to lead this effort would be the Economic Development Administration, in the U.S. Department of Commerce. Given EDA's limited funding, it might be more effective for EDA to focus on improving the economic development programs of state and local areas, rather than on carrying out its own programs.

(2) <u>Greater professionalization of economic development agencies and legislative review committees</u>. Over time, the economic development profession has become more professionalized.

This trend needs to be continued and the idea of high-quality evaluation as an essential part of carrying out high-quality economic development programs should be encouraged.

(3) Demonstration effect of good evaluations. If some good evaluations are done, this should lead to more good evaluations. It is difficult to get people to do something that hasn't been done before. Once state and local economic development policymakers have seen that a high quality evaluation of economic development programs can help improve the program's performance and political viability, the interest in economic development evaluations should increase. While much of the discussion here has tended to emphasize outcome evaluations (because they are so infrequent) in economic development, good practices would suggest that process and outcome evaluations be accomplished simultaneously. After all, a program can hardly be shown to accomplish what it intends if it is not properly implemented.

Developing a tradition of high quality evaluations of economic development programs is likely to take some time. Economic development evaluation is where job training evaluation was 20 years ago—with a few good evaluations, more low quality evaluations, and too few evaluations overall. Over the next 20 years, enough high quality evaluations of economic development programs should be done so that development professionals really know what works and what doesn't work in economic development.

References

Bartik, T.J. (1993, September). "Who Benefits From Local Job Growth, Migrants or the Original Residents?" *Regional Studies*, Vol. 27, No. 4 : (pp. 297-311).

______________. (1992, February). "The Effects of State and Local Taxes on Economic Development: A Review of Recent Research." *Economic Development Quarterly* : (pp. 102-10).

___________. (1991). *Who Benefits From State and Local Economic Development Policies?* Kalamazoo, MI: W. E. Upjohn Institute for Employment Research.

Benus, J.M., Wood, M., and Grover, N. (1994, January). *A Comparative Analysis of the Washington and Massachusetts UI Self-Employment Demonstrations*. Report prepared for the U.S. Department of Labor, Employment and Training Administration, Unemployment Insurance Service under Contract No. 99-8-0803-98-047-01.

Bingham, R.D. and Felbinger, C.L. (1989). *Evaluation in Practice: A Methodological Approach*. New York: Longman.

Burtless, G. and Haveman, R.H. (1984). "Policy Lessons from Three Labor Market Experiments." In R. Thayne Robson (ed.) *Employment and Training R&D*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, (p. 128)

Commonwealth of Virginia. (1991). *Review of Economic Development in Virginia*. Report of the Joint Legislative Audit and Review Commission, The Virginia General Assembly, House Document No. 39.

Elder, A. H. and Lind, N.S. (1987, February). "The Implications of Uncertainty in Economic Devleopment": The Case of Diamond Star Motors." *Economic Development Quarterly:* Vol 1 No. 1, (pp. 30-40).

Hatry, H.P., Winnie, R.E. and Fisk, D.M. (1981). *Practical Program Evaluation for State and Local Governments*. 2nd. ed., Washington, D.C.: Urban Institute.

Holzer, H.J., Block, R.N., Cheatham, M., and Knott, J.H. (1993, July). "Are Training Subsidies for Firms Effective? The Michigan Experience." *Industrial and Labor Relations Review*, Vol. 46, No. 4 : (pp. 625-36).

Illinois State, Office of the Auditor General. (1989, July). *Management and Program Audit of the Department of Commerce and Community Affairs Economic Development Programs*. Report prepared for the Legislative Audit Commission.

Jencks, C. (Winter 1992-1993). *Focus,* University of Wisconsin-Madison, Institute for Research on Poverty : (p. 12).

Jones, B.D. and Bachelor, L.W. (1984). Local Policy Discretion and the Corporate Surplus." In Bingham, R.D. and Blair, J.P. (eds.), *Urban Economic Development*. Beverly Hill, CA: Sage, (pp. 245-267).

Knowledge Systems and Research, Inc. (1991, August). *Evaluation of Twelve New York State Technical Assistance and Financing Programs*. Report prepared for New York State Department of Economic Development and New York State Urban Development Corporation.

Markusen, A.R. (1985). *Profit Cycles, Oligopoly, and Regional Development*. Cambridge, MA: MIT Press.

Mt. Auburn Associates. (1992, December). *An Evaluation of Ohio's Thomas Edison Technology Centers*. Final Report submitted to The Ohio Department of Development.

__________ (with the Corporation for Enterprise Development). (1990, June). *The Economic Impact of Business Financing by New York State Agencies*. Report prepared for the Department of Economic Development, Urban Development Corporation, Job Development Authority, and Science and Technology Foundation.

Papke, L.E. (1991, December). "Tax Policy and Urban Development: Evidence from an Enterprise Zone Program." NBER Working Paper No. 3945.

Pennsylvania Economy League, Inc. (1992, May). *Impact of Northeast Tier, Ben Franklin Technology Center on Its Region*. Report prepared for Northeast Tier, Ben Franklin Technology Center.

Persky, J., Felsenstein, D. and Wiewel, W. (1993, July). "A Methodology for Measuring the Benefits and Costs of Economic Development Programs: Technical Documentation." Chicago, IL: University of Illinois at Chicago/Center for Urban Economic Development.

Price Waterhouse. (1992, March). *Evaluation of the 7(a) Guaranteed Business Loan Program*. Final Report prepared for the U.S. Small Business Administration under Contract Number SBA-5033-FAD-90.

Public Policy Associates and Brandon Roberts and Associates. (1992, October). *Oregon Small Business Services Evaluation*.

Rephann, T.J. (1992). "Highways and Rural Economic Development: Evidence from Quasi-Experimental Control Group Studies." West Virginia University Regional Research Institute Research Paper 9210.

Rohde, S., Cash, J., and Ammarman, K. (1990, March). "Study of the Capital Access Program" Michigan Strategic Fund working paper.

Rossi, P.H. and Freeman, H.E. (1985). *Evaluation: A Systematic Approach*. *3rd ed*. Beverly Hills, CA: Sage.

Trisko, K.S. and V.C. League. (1978). *Developing Successful Programs*. Oakland, CA: Awareness House.

Wiewel, W., Persky, J. and Felsenstein, D. (forthcoming 1994). "The Costs and Benefits of Business Subsidies." *Economic Development Commentary*.