This chapter narrows the discussion of the methods described in the last chapter to the field of early childhood intervention programs. Following the framework of the scorecard presented in Chapter Two and in particular the row elements, we first outline some of the important issues related to measuring costs for these programs. Next, we describe the outcome domains that are relevant to early childhood programs, and how those outcomes are translated into program benefits (or costs avoided). The chapter closes with a discussion of specific issues in aggregation associated with comparing benefits and costs.

MEASURING COSTS OF EARLY CHILDHOOD INTERVENTION PROGRAMS

As discussed by Karoly et al. (1998), an extensive literature evaluates the impacts of early childhood intervention programs for participating children and their families. While evaluations of early childhood intervention programs have led to an established base of research focused on outcomes, there is less of a basis for assessing program costs. Information about program costs is often not reported in the evaluation literature and may not even be collected during the course of a demonstration project or larger-scale evaluation. As the discussion in Chapter Two conveys, however, cost information is an essential component of the types of cost and outcome analyses available to inform investment decisions in social service programs. Cost information may be used to evaluate the benefits versus costs to
society of a given program or to guide innovation and improvement in program design.

Regardless of the goal of the cost and outcome analysis, the operational aspect of assembling the cost data is tantamount to filling in the cost element panel of the scorecard (Table 2.1). For each column in the scorecard, cost elements should be broken out by who bears various costs, when costs are incurred, and other aspects that would vary depending on the goal of the analysis. In the remainder of this section, we first discuss some of the general principles that guide cost measurement. We then discuss some of the details regarding the row elements and provide a hypothetical example of the cost elements in a scorecard.

Some General Principles of Cost Measurement

In measuring the costs of a program for the purposes of policy scorecard analysis, the goal is to enumerate the comprehensive set of resources forgone by all parties who might incur some loss as a result of the program. That is, the costs of a program are not entirely captured by the budget an agency uses to fund the program. Rather, a more comprehensive characterization of the costs of a program would capture the difference in resources required for a world without the program (the baseline) and the same world with the program, as discussed earlier in Chapter Two. This broader notion of costs allows for the fact that entities other than the agency—such as program participants and other members of society—might also incur some costs in a world with the program. It also recognizes that not all costs involve explicit expenses, but rather that some costs might take the form of in-kind resources devoted to the project, such as volunteer time or subsidized facilities.

This construct—program cost as the difference between total costs in a world with and without the program—highlights the importance of having a control (or comparison) group. Cost data from the control group serve as estimates of the costs of the world without the program, and data from the intervention group serve as estimates of the cost of the world with the program. If no cost data from a control group are comparable to the cost data from the intervention group, the estimates of the costs of the program will be fraught with considerably more uncertainty and error.
When collecting cost information for the control group, it is important that information be gathered for the same set of service providers as for the intervention group. This is essential to capture possible cost shifting (e.g., from one service provider or payor to another) or cost offsets (e.g., reduced use of services in one area as a result of increased services use in another). If cost information is more narrowly collected for the control group, it is possible to miss changes in the mix of services used or the total amount of services used as a result of the program (see, e.g., Foster and Bickman, 2000).

As discussed in Chapter Two, the objectives of the analysis dictate some of the particulars of cost estimation. For example, if the overarching goal is to compare the benefits and the costs of the program, then it is enough to estimate a single number or range of numbers (e.g., the cost is between $1.1 million and $1.3 million). However, suppose the objective is to estimate the cost of a similar program implemented somewhere else, or to use the cost estimates to guide a continuous quality improvement (CQI) effort. In these cases, it would be more useful to develop cost estimating relations (CERs), which estimate cost elements as a function of the design of the program. These relations generate various cost elements as a function of design variables, such as types of personnel who provide services, intensity of treatment, equipment and facilities required, and other potentially variable features of the program.

**Types of Costs**

There are various ways to categorize resources, but here we focus on some of the major categories that are likely to be particularly salient for early childhood programs. These categories help ensure comprehensive accounting of all resources that a program requires.

Cost analysts frequently categorize resources associated with program delivery into personnel, equipment, facilities, and supplies/other. Personnel includes all labor, e.g., social workers, nurses, secretaries, drivers, maintenance workers, and administrative personnel. Equipment includes durable items, such as office equipment (copiers, printers, computers, desks) and vehicles (automobiles, buses). Facilities includes land, office space, garage space, parking space, and maintenance sheds. Supplies/other includes consumable items, such as paper and ink for copiers and printers, gasoline for the
vehicles, and coffee for the personnel. Utilities can be included in this category or broken out separately. In any particular study, if a category is too small (e.g., less than 5 percent of the total), the cost analyst may combine it with another. If a category is too large (e.g., more than 40 percent of the total), the cost analyst will split it into subcategories.

An important distinction in costs is between explicit expenses and in-kind resources. Obviously, costs that are billed need to be counted. It is also important to capture costs that accrue in the course of providing services but do not involve a monetary transfer. These likely will involve in-kind resources provided to the program from outside the agency, such as subsidized rent for facilities or meals provided by other government agencies.

Who pays for a resource is important. Cost analysts typically distinguish between internal and external costs. Internal costs fall on the agency that sponsors the program. External costs fall elsewhere. However, this distinction is often inadequate. Instead of distinguishing only between internal and external costs, one should distinguish costs (and benefits) by stakeholder. If there are a dozen stakeholders, there should be a dozen “who pays” categories.

For example, participants may bear certain costs to participate in the program. These would include the costs of transportation to appointments or lost wages from missed work. In the case of early childhood programs, it is especially relevant to consider costs borne not only by participating children, but also by their parents or caregivers, even when the latter group is not explicitly a focal point of the treatment program.

Another example of a cost that the agency providing services does not bear is the costs generated by referrals to other services. This is sometimes referred to as cost shifting and is important to capture in programs designed to increase use of other services. (Use of other services by providers outside the intervention may also decline.) For example, in the Elmira Prenatal/Early Infancy Project (PEIP)—a nurse home visiting intervention discussed more fully in Chapter Four (Olds et al., 1997)—part of the treatment provided by home visitors was to refer participants to other social services for which they might qualify. While greater use of these other social services did not
impose a cost on the PEIP, it clearly raised costs for the other agencies that provided the additional services.

Collecting cost data for the control group (or baseline) for the same set of service providers as for the treatment group allows such cost shifting to be detected, though the analyst must give a priori thought to where cost shifting may occur and be sure to measure it. A comparison of cost data for the control group versus the treatment group will also reveal any cost offsets, whereby costs are reduced for services outside of the treatment program that are used by program participants.

Another way to categorize resources is to distinguish between consumable and nonconsumable items. A consumable item—such as paper or gasoline—is measured in units of quantity, such as reams or gallons. Nonconsumable items—such as facility space, durable goods, and personnel—are measured in units of quantity used per time unit—e.g., square-feet-months or person-years.

It is also frequently useful to distinguish between fixed and variable costs. Fixed costs are likely to be onetime costs, which often occur early in implementing a program. Examples of fixed costs are the costs of developing a curriculum or treatment protocol, and the costs of constructing facilities when they are not rented. The key feature of fixed costs is that they do not vary with the amount of time the program is in place. Variable costs are those that accrue in each time period the program operates, such as utility bills and payments to staff.

Cost analysts also distinguish between investment costs and operating costs. Investment costs are sometimes called nonrecurring costs. They are incurred to start a program or to increase its scale.\(^1\) Often they pay for increases in nonconsumable resources such as vehicles or facilities. Operating costs are recurring costs; they must be paid each year to keep the program running. They are often assumed to be proportional to the inventories of nonconsumable resources on hand (e.g., salary plus benefits of an employee) or to the annual quantity of a resource consumed (the constants of proportionality

---

\(^1\) In some studies it is useful to split nonrecurring costs into research and development (R&D) costs and investment costs.
are often called cost factors. Costs that have already occurred (or have already been contracted for) and cannot be recovered are sunk costs. These can correspond to resources that are on hand and cannot be sold, or in-kind contributions (e.g., volunteer labor, office space) which will only become available if the program is implemented. They should not influence one’s decision whether to invest in the program, because they will be the same even if one does not invest.2

Another noteworthy feature of costs is that they accrue over time and are likely to display variation over time. For instance, program costs might be high at the time of inception as the fixed costs of setting up facilities and training staff are born. Program costs might drop during a period when participants are screened or diagnosed, and then rise again during a “treatment” phase. It is useful to construct a variant of the scorecard whose rows are resource categories and whose columns are years. For consumable resources (those in the supplier/other category), each cell contains the amount of the resource consumed in that year of the program’s operation. For nonconsumable resources, each cell contains the inventory of the resource on hand at the end of that year. This table is easier to construct than it may seem. Typically a program will start small and build capacity over time. So the analyst determines the resources needed by the mature program (say, in year five and beyond), and ramps up the resources over years one through four to achieve those levels.

In sum, cost accounting entails a number of categorizations. In filling in the cost elements of the scorecard, the analyst considers the various stakeholders, how costs accrue over time, whether the costs are in-kind or explicit, as well as a number of other considerations.

Capturing the Sources of Cost Variation

There might not be one simple answer to the question, “How much does this program cost?” The answer may depend on a number of factors, such as:

2There is an unfortunate temptation to let sunk costs affect one’s decisions. “I’m going to hang onto that stock until it gets back to the price I paid for it.” It is more profitable to base a decision to buy or hold or sell a stock on its future performance, not its past performance.
• From whose perspective are costs calculated—the participants, the government, society as a whole, etc.?
• In what geographic location or site is the program?
• What are specific design features of the program?

Capturing who pays for a particular resource enables one to calculate costs from different points of view. Social programs have many stakeholders, and a particular program may provide net benefits to some stakeholders while extracting net costs from others. It is not enough to calculate total net costs (and total net benefits) if the parties who pay the costs do not reap the benefits. A corollary to this is that costs may appear in more than one place in the cost model. That is, a cost element may be a cost to one party and a benefit to another, and hence will appear twice in the model (with opposite signs). When aggregated to society’s perspective, these two would cancel out.

Even when program protocols are followed uniformly across locations, the program costs will likely vary by geographic location or site. This is because differences in costs result from such factors as the cost of rental space in a local area, whether a site is in an urban or rural area, and the relative wage rates of staff. For example, the transportation costs in rural areas, which often lack low-cost public transportation systems, might be considerably higher than those in urban areas. On the other hand, wages and rental prices are often lower in rural areas than in urban areas.

Another source of cost variation is the design features of the program. One’s first inclination may be to take measures of the program’s workload as the design variables. For an early childhood program, this might be the number of participants. However, additional design variables (e.g., capacity) are likely to be needed to portray costs fully. Often a program will be designed to have a given capacity, and some costs will be incurred whether or not the capacity is actually used. For instance, a group meeting requires a program staff member to be present no matter how well or poorly attended it may be. Omitting the capacity variables amounts to making an assumption about the utilization rate, which may have a strong influence on costs.
Another type of design feature that would generate variation in costs is the risk category of participants. Participants in different risk categories might utilize different types of services or require different lengths or intensities of treatment. Risk category may also yield differences in outcomes and hence benefits, and as a result the entire analysis could hinge on the distribution of participants across risk categories (Karoly et al., 1998).

Numerous additional design features might contribute to variability in program costs, such as whether the staff are medical doctors versus registered nurses or whether the participants are treated in a group or individual setting. While a complete list of design features would be too numerous to describe here, this discussion has suggested types of issues that need to be considered.

A Brief Hypothetical Example of Cost Elements and Data

To help fix the ideas we have discussed related to costs, we provide a brief example of some hypothetical cost elements and data in Table 3.1. Table 3.1 assumes a baseline and a new program, each of which provides some type of service to parents of young children. The cost elements and their values are completely fictional, but are a realistic representation of potential stakeholders, types of cost elements, and units of measure. In this table, we present hypothetical costs for the baseline program versus hypothetical costs for the new program. These are indicated by columns 2 and 3 in the table. The cost of the new program in this context—column 4—is the difference in costs between columns 3 and 2. That is, column 4 shows the incremental costs of the program over and above the baseline.

Cost elements are chosen to capture the resources employed by the program and who pays. We have chosen four hypothetical sets of stakeholders for whom there will be cost elements in this example: participants, the agency implementing the program, other agencies (that might provide services to which participants are referred), and the rest of society. We have indicated specific cost elements only for participants. These cost elements should represent all explicit, in-kind, and implicit resources the participants would incur when participating in the baseline program and the new program. These
### Table 3.1

**Hypothetical Examples of Cost Elements for Baseline Program and New Program**

<table>
<thead>
<tr>
<th>Hypothetical Cost Elements</th>
<th>Baseline Program (2)</th>
<th>New Program (3)</th>
<th>Difference Between (3) and (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of visits per year</td>
<td>4</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Time per visit</td>
<td>60 minutes</td>
<td>80 minutes</td>
<td>20 minutes</td>
</tr>
<tr>
<td>Wages per hour</td>
<td>$6.25</td>
<td>$6.25</td>
<td>0</td>
</tr>
<tr>
<td>Miles per trip to visit</td>
<td>15</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>Cost per mile</td>
<td>$0.20</td>
<td>$0.20</td>
<td>0</td>
</tr>
<tr>
<td>Copayments per visit</td>
<td>$5</td>
<td>$5</td>
<td>0</td>
</tr>
<tr>
<td>Paperwork time per year</td>
<td>35 minutes</td>
<td>50 minutes</td>
<td>15 minutes</td>
</tr>
<tr>
<td>Child care hours per year</td>
<td>0</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Prescriptions filled per year</td>
<td>3</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

Agency Implementing Program
Other Agencies
Society as a Whole

might include time expended and explicit cash outlays. Examples of cost elements in the table that might be assigned to participants are the time length of the visits, number of visits, visit copayments, paperwork time, number of prescriptions, and hours of child care (primarily when a parent is at additional appointments or meetings resulting from referrals).

It is important to measure cost elements in terms of resources, and only later price out the resources to obtain dollars. Costs should not be expressed directly in dollars, unless the resource inventories behind the dollars are unavailable. As was mentioned in Chapter Two, there are several reasons for this. These include variation in prices across locations, avoiding accounting assumptions regarding shared resources, and the need to substitute for resources not available in another site.

In Table 3.1, this is demonstrated in several cost elements. Transportation, for instance, is expressed in the number of trips rather than dollars, and the miles per trip and cost per mile are indicated in separate cost elements. This way, sites where participants use differ-
ent modes of transportation, such as subway, bus, cab, or the participant’s car, can account for the differences in time costs and cash outlays inherent in those modes of transportation. Similarly, the time per visit is expressed in minutes. Sites that serve participants who work would value this time differently from sites whose participants are largely out of the labor force.

As discussed in Chapter Two, scorecard entries will ideally include information that characterized their statistical uncertainty. While not shown explicitly in this hypothetical table, cost ranges or confidence intervals could be included as well as expected values. The degree of statistical uncertainty surrounding cost estimates would be important to consider when comparing costs across programs.

As a final point in discussing costs, note that gathering data for the analysis described here is itself a cost. In addition to the resources consumed by the analysis team, note that the analysis is likely to impose costs on the program itself. The program staff likely will be required to provide or collect data, which will require additional time, training, and perhaps even computers or other equipment.

OUTCOME DOMAINS AND MEASURING BENEFITS

As noted above, there is a long history and well-developed methodology for measuring the impacts of early childhood intervention programs on participating children and their families. Since the 1960s, a wide array of smaller- and larger-scale early intervention programs have been implemented and formally evaluated, often with experimental designs to allow comparison of outcomes for program participants versus a randomly assigned control group. While the randomized control trial remains the gold standard for evaluating social service delivery programs, some evaluations adopt quasiexperimental designs using matched comparison groups as controls. The experimental and quasiexperimental early intervention evaluation literature has been synthesized in a number of comprehensive reviews. For recent examples, see Barnett (1995), Yoshikawa (1995), Guralnick (1997), Reynolds et al. (1997), Karoly et al. (1998), and Currie (forthcoming). Regression analysis and related methods are another set of tools that can provide insights into service delivery questions, such as whether a program improves outcomes of participants or whether different populations realize different outcomes from a program (see, for example, Currie and Thomas, 1995, or NICHD Early Child Care Research Network, 1997). Also, see Hargreaves et al. (1998), Chapter 9, for a discussion of the use of regression techniques in cost and outcome methods.
tions generate specific measures of program impacts for particular individuals at given points in time, either when program services are delivered or after the intervention program has ended.

Just as program costs are measured net of a baseline without the program being evaluated, the outcomes of early intervention programs are net impacts (i.e., the same outcomes are measured for both the treatment and control/comparison groups for the same period of follow-up, and the program effects measure the difference between the two groups). Typically, these program impacts are measured in quantities other than those denominated in dollars. The challenge for cost and benefit analysis is to translate the beneficial effects of early intervention programs measured in such units as IQ points, years of special education, months employed, or counts of juvenile crimes into dollar values that can then be compared with program costs. The remainder of this section considers the types of program impacts typically included in evaluations of early intervention programs and the approaches available for translating these outcomes into dollar benefits.

**Measuring the Impact of Early Childhood Intervention Programs**

Targeted early intervention programs can be viewed as sharing a common aim: to improve child health and development by providing socioeconomically disadvantaged children and their families with various services and social supports during part or all of the period of early childhood (Karoly et al., 1998). Despite this common aim, considerable variation occurs in early intervention program objectives and designs and in the associated services and supports provided to meet the program goals. Likewise, program evaluations are not uniform in the outcome measures collected. Instead, resource constraints for data collection and other factors limit most evaluations to capturing only a subset of measures that reflect the domains where the program is expected to have an impact, whether for the focal child or for their parents and other caregivers.4

---

4While program evaluations do not always collect the same set of outcome measures, for those measures conceptually similar to those collected in other evaluations, it is often desirable to use common measures so that comparisons can be made. For
Table 3.2 illustrates the range of measures of early intervention program impacts in four broad domains: emotional and cognitive development, education, economic well-being (e.g., public assistance receipt, income, crime), and health. Within each domain, we list some of the most frequently used measures in early intervention studies, either for participating children or for their parents and other caregivers. (In support of the discussion in the next subsection, italics are used in the table to indicate which of these outcomes are most readily translated into dollar values.) The specific measures in Table 3.2 are intended to illustrate the types of outcomes measured in each domain, rather than reflecting the full range of measures used in the evaluation literature. We discuss each of these domains in turn, as well as some more general measurement issues common across domains.

**Emotional and Cognitive Development.** Given the goal of early intervention to enhance child development, most early intervention evaluations include measures in this domain, either for participating children or their parents and other caregivers. For children, the measures include scores on batteries that measure socioemotional development or behavioral problems, as well as cognitive development—typically IQ scores. For parents, scales are used to measure aspects of the parent’s role in the child’s development, such as the nature of the parent-child relationship and the quality of the home environment. The specific scales and tests used are selected to be age-appropriate (whether administered to a parent or child) and to reflect the specific objectives of the program being studied. To select measures that are reliable and valid, with well-known psychometric properties, many interventions often use the same specific scales or test batteries, such as the Stanford-Binet or Wechsler intelligence tests to measure IQ, or the HOME Inventory to assess parental caregiving and the home environment.
Table 3.2

Early Childhood Intervention Program Outcome Domains and Illustrative Measures

<table>
<thead>
<tr>
<th>Outcome Domain</th>
<th>Illustrative Measures for:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Child</td>
</tr>
<tr>
<td>Emotional and cognitive development</td>
<td>Socioemotional and</td>
</tr>
<tr>
<td></td>
<td>behavioral scores</td>
</tr>
<tr>
<td></td>
<td>IQ test scores</td>
</tr>
<tr>
<td></td>
<td>Teacher’s ratings</td>
</tr>
<tr>
<td>Education</td>
<td>Achievement test scores</td>
</tr>
<tr>
<td></td>
<td>Grades</td>
</tr>
<tr>
<td></td>
<td>Grade progression</td>
</tr>
<tr>
<td></td>
<td>(repetition)</td>
</tr>
<tr>
<td></td>
<td>Participation in special</td>
</tr>
<tr>
<td></td>
<td>education</td>
</tr>
<tr>
<td></td>
<td>Educational attainment</td>
</tr>
<tr>
<td>Public assistance receipt, income,</td>
<td>Receipt of public assistance</td>
</tr>
<tr>
<td>crime</td>
<td>Employment</td>
</tr>
<tr>
<td></td>
<td>Earnings/income</td>
</tr>
<tr>
<td></td>
<td>Criminal activity</td>
</tr>
<tr>
<td></td>
<td>Contact with criminal</td>
</tr>
<tr>
<td></td>
<td>justice system</td>
</tr>
<tr>
<td>Health</td>
<td>Physical and mental</td>
</tr>
<tr>
<td></td>
<td>health status</td>
</tr>
<tr>
<td></td>
<td>Child abuse and neglect</td>
</tr>
<tr>
<td></td>
<td>Substance abuse</td>
</tr>
<tr>
<td></td>
<td>Fertility control</td>
</tr>
<tr>
<td></td>
<td>Emergency room visits</td>
</tr>
<tr>
<td></td>
<td>Other health care use</td>
</tr>
</tbody>
</table>

NOTE: Italics indicate measures more easily expressed in dollar terms.

**Education.** Another common aim of early intervention programs is to improve school readiness and subsequent school performance. Consequently, a great deal of interest has arisen in tracking educational outcomes for program participants versus those in the control or comparison group. Prior to school entry, few direct measures of school readiness exist although researchers often consider measures of cognitive development and socioemotional regulation and control as relevant indicators. For school-age children, evaluations typically measure scores on achievement tests in reading, math, or other subjects. Achievement test scores at older ages are included in longer-
term evaluations as well. Longer-term follow-up also allows measurement of educational outcomes relevant at older ages, such as grade progression (or alternatively grade repetition), use of special education, high school completion, and eventual educational attainment. In some cases, early intervention programs are designed to improve educational outcomes for parents as well, so educational attainment is also measured for them.

**Economic Well-Being.** Early intervention programs may also affect other areas of functioning during adolescence and adulthood. If early intervention programs improve socioemotional development and educational performance, those gains may translate into improved economic well-being. With longer-term follow-up, for example, some programs have been evaluated in terms of their impact on economic outcomes such as dependence on social welfare programs (e.g., use of public assistance or “welfare,” Food Stamps, or Medicaid) and labor market performance or economic success (e.g., employment rates, occupational status, earnings, income, poverty status). Another area of assessment is involvement in criminal activity, either by directly measuring specific crimes committed or by quantifying contact with the criminal justice system (e.g., arrests, convictions). While program evaluations typically consider these outcomes for participating children as they make the transition to adulthood, some programs have assessed parents and other caregivers in this domain using similar measures.

**Health.** This final category captures the expectation that early intervention programs may affect health outcomes, broadly defined to include aspects of health status and health care use. In addition to evaluating the impact of early intervention on general physical or mental health status, some programs consider more specific areas of health, such as the incidence of child abuse and neglect, perceived quality of life, family violence and substance abuse, impairment, and fertility control (e.g., the timing and spacing of births). Health care use may also be affected by early intervention programs, with some programs focusing on costly emergency room visits, as well as other forms of health care use (e.g., hospitalizations or use of specific health care services). While many program evaluations focus on these measures for participating children, either at younger or older ages, these measures may also be assessed for parents and other
caregivers when early intervention services are designed to affect their health status or health care use.

Across these four domains, a number of common measurement issues arise. A first concern is whether to focus outcome measurement on the participating child or the parent and other caregivers, and for each of these potential beneficiaries of program services, whether the intervention impacts can be captured in the short run or the long run. As indicated in Table 3.2, early intervention programs may benefit not only the children they serve but also their parents or caregivers. The first generation of early intervention programs and their associated evaluations focused on child outcomes (see, for example, the studies cited in Karoly et al., 1998). With a growing recognition of the importance of the family and home environment in the early years of life and of the potential for programs to impact parental outcomes, program services and evaluations have incorporated the parental side of the equation as well. If a program can be expected to affect parental outcomes, many of those outcomes listed in Table 3.2 can be captured in a short-term evaluation, as they may be measured during the period of service delivery or soon after the program ends. In contrast, many of the outcomes in Table 3.2 listed for children cannot be directly assessed without follow-up that extends many years, if not multiple decades, beyond the period of program delivery. Such long-term follow-up requires a significant commitment of resources to execute as well as to minimize the biases associated with attrition in longitudinal studies.6

Another methodological concern is whether measures should capture contemporaneous outcomes or a longer history of a given outcome. This is particularly relevant for evaluations that include long-term follow-up. Consider the case of public assistance utilization, either by the participating child’s family during childhood or by the child when the child reaches adulthood and forms a household of his or her own. During any given assessment, either during the intervention or in a subsequent follow-up, it is possible to collect information on current program utilization (i.e., whether the individual is

6It is important to standardize the period of follow-up or future projection if programs are to be strictly compared in terms of costs and benefits. Otherwise, the program with a longer follow-up or with projections further into the future will likely be favored on cost-benefit terms.
currently receiving support). An alternative is to collect data on a partial or complete history of program participation during the interval between the present and the last point of data collection (or even prior to the baseline). While the intervening history clearly requires more effort to collect, it provides the information necessary to value a continuous sequence of potential benefits. A reduction in public assistance utilization in each year of a 10-year horizon will clearly translate into greater savings to government than will a reduction in utilization for the final year of that horizon (e.g., the tenth year). This consideration is relevant for many of the outcomes listed in Table 3.2, including measures of educational outcomes (e.g., grade repetition, special education use), economic outcomes in addition to use of social welfare programs (e.g., employment, earnings, income, criminal activity, and criminal justice system contact), and health outcomes (e.g., health care utilization).

A final measurement issue is the method of collecting the specific outcome indicators. During the period of program intervention, the measures listed in Table 3.2 (and others in the four domains not listed) are typically collected through some form of interaction with the study participants (those receiving the treatment as well as controls). Survey questionnaires, test batteries, direct observations, and program administrative records may be appropriate depending on the specific outcome of interest. Once the intervention has ended, continued assessment may require continued personal interaction with study participants or possibly the reliance on external sources of information, such as administrative records. For example, with the proper human subjects consent procedures, information on criminal activity (e.g., arrests, incarcerations) may be collected by interviewing participants during a follow-up or by tracking activity recorded by the criminal justice system. Administrative data can be useful for collecting information on other outcomes, such as school performance, participation in social welfare programs, and employment outcomes.

Administrative data have several advantages. They may be free of various reporting biases and may result in lower rates of missing data (or cases lost due to nonresponse). This is especially true for longer-term follow-up when respondents may have difficulty with long-term recall of specific events (e.g., a monthly employment history) or may not be even available for an interview because they cannot be
located. However, tracking outcomes through administrative sources requires advanced planning to secure the necessary permissions from study participants. Administrative data are often not released due to concerns about protecting individual privacy, and individuals may still be lost to follow-up when they cannot be tracked across administrative boundaries (e.g., state borders).

Translating Program Impacts into Dollar Benefits

Once a formal program evaluation has measured the impact of an early intervention program using one or more of the measures listed in Table 3.2, many of the analysis methods reviewed in Chapter Two require that the analyst convert that outcome to a monetary value. The process of expressing the benefits in dollar terms, or “monetizing” the program impacts, is easier for some of the outcomes listed in Table 3.2 than for others. This reality is illustrated in Table 3.2 by denoting those outcomes that are most readily monetized in italics. Those outcomes not in italics may still be expressed in dollar terms but only with less reliable benefit-cost estimates or by virtue of more heroic assumptions.

The economic outcomes listed in Table 3.2 are among those that are the easiest to monetize, whether the program impact is lower public assistance utilization and the benefit is reduced outlays by local, state, or federal governments or the program impact is more months spent employed and the benefit is higher taxes paid. To illustrate, consider an early intervention program evaluation, which shows that at the age 15 follow-up, the families of children who participated in the program used 20 fewer months of public assistance benefits over the past 10 years than did families in the control group. If each month of benefits costs taxpayers $500 (including both cash payments and administrative costs), this early intervention program would lead to dollar savings to government of $10,000. After spreading those savings over each relevant age (from six to 15) and

---

7These are benefits from the point of view of the government, and we adopt this point of view for illustration only. The analyst should be prepared to present costs and benefits from the point of view of any stakeholder. For example, from the point of view of society as a whole, taxes are not a benefit but a transfer payment, and one should use incremental income instead.
discounting to a specific time period (e.g., birth) using a specific discount rate (e.g., 4 percent), the NPV of the savings to government could be calculated ($6,410 in this case). A similar process can be followed for each of the italicized outcomes shown in Table 3.2.8

Other outcomes listed in Table 3.2 cannot be translated into dollar values with the same ease. For example, many early intervention programs demonstrate short-term and long-term gains in cognitive measures, such as IQ or achievement test scores. This impact is difficult to translate into a dollar value. However, if these cognitive benefits lead to improved educational and economic outcomes, then valuation of outcomes in these collateral domains may capture, at least in part, some of the benefits of better cognitive outcomes.

The process of assigning an economic value to a given program impact is not always uncontroversial. One outcome that invites differences of opinion is the value to society of the reduction in criminal activity stemming from early intervention. As discussed in Chapter Two, while most experts agree on the value of the tangible costs associated with criminal acts based on empirical evidence (e.g., costs for property loss, medical expenses, lost income due to injury), there is less agreement over the value to assign the intangible costs (e.g., pain and suffering of crime victims). Different methods of valuing pain and suffering can lead to widely different estimates of the intangible costs of crime. For instance, Klaus (1994) estimates the cost of a rape to be $234, whereas Miller et al. (1996) use a figure of $5,100. Based on personal experience, some audiences believe a figure in the $5,000 range is much too low. This type of controversy may affect other areas of program impacts, particularly when empirical evidence regarding economic values is weak or nonexistent.

In some cases, it is possible to assign benefits beyond the period of direct observation. For example, improvements in educational attainment can be associated with an entire earnings profile from young adulthood to age 65 based on other studies of earnings trajectories in the literature (for an example, see Barnett, 1993). On the basis of criminal activity through adolescence or early adulthood, the individual’s future criminal “career” in adulthood can be forecast

---

8For more detail on these types of calculations, see the cost-benefit studies cited in Chapter Four.
(see Karoly et al., 1998, for an example). Although such forecasts introduce additional uncertainty into the benefit calculations, they do help overcome the limits of follow-up periods that end in early adulthood when the economic benefits for participants in early intervention programs may just be beginning to be realized.

When intermediate impacts (e.g., educational attainment) are used to value longer-term impacts (e.g., earnings), it is important to avoid double counting program benefits. In some cases, the intermediate outcome may generate benefits in and of itself, in addition to providing information to project benefits for a longer-term but unobserved outcome. In the case of educational attainment, if an early childhood program increases years of schooling for the treatment group compared with the control group, educational costs actually increase because of the additional time spent in school. At the same time, the higher educational attainment can be used to project earnings gains throughout adulthood compared with the trajectory that would be experienced with a lower level of attainment. However, if actual earnings are observed for any period beyond the intervention, the projected earnings should not be counted for the same age span.

**COMPARING COSTS AND BENEFITS**

In comparing the costs and benefits of an early childhood intervention program, two critical issues are the following:

- Who pays the costs versus who realizes the benefits?
- What is the decision rule for selecting the best alternative?

These two issues are related. We discussed the first issue earlier in pointing out that various costs are borne by different stakeholders, such as children, parents, government, and society as a whole. The benefits could be realized by one party—such as the children—while the costs are paid by another party—say, the government.

The second issue helps resolve this quandary. As discussed in Chapter Two, various decision rules might be specified that would yield different answers to the question of whether the benefits outweigh the costs and for whom. In cost-savings analysis, the costs of a program to government are compared to the savings of a program to
government. If the latter outweigh the former, then the program pays for itself from the perspective of government. In cost-benefit analysis, the costs of the program borne by all of society—including participants, government, and others—are compared to the total benefits accrued to any of the parties. This calculus is indifferent to who pays and who benefits.

Chapter Two noted some specific methodological issues associated with these various approaches, including choosing a discount rate, accounting for scenario uncertainty, and capturing statistical uncertainty. An additional challenge in the comparison of costs and benefits likely to be particularly relevant for early childhood intervention programs is the fact that they may accumulate at different rates. These programs typically intervene briefly in the early years of a child’s life. In contrast, the benefits may take years to accumulate, as the child’s outcomes in such areas as high school graduation, adult employment, and public assistance participation become apparent. This creates a potential temporal mismatch between the payment of costs and the realization of benefits, even if the measure of merit considers only costs and benefits to government. This is because the government—i.e., the taxpayers—that pays for the program might not be the same government (taxpayers) that reaps the benefits two decades later when the treated children enter adulthood.

As discussed further in the next chapter, Karoly et al. (1998) demonstrate that the costs of the Perry Preschool program take two years to accumulate compared to the benefits, which accumulate to the level of costs after nearly two decades. In contrast, another program reviewed in the next chapter, the nurse-home visiting model known as the Elmira Prenatal/Early Infancy Project, generates benefits earlier in the life course because of changes in the parents’ behavior (specifically, the mother). In that case, the benefits accumulate more rapidly and are realized at a level that exceeds program costs shortly after the two-year intervention ends.

Another challenge for these tools is the conservative nature of most estimates of program benefits. Due to the limitations of placing an economic value on the benefits of early intervention, most cost-benefit studies of these programs are likely to understate the benefit
side of the equation for two reasons. First, many of the benefits of early intervention programs may not even be measured as part of the evaluation. This may stem from resource constraints that limit the number of measures collected or because some measures are more difficult to collect. For example, some early intervention programs may produce spillover benefits for other siblings (e.g., as a result of improved parenting or better economic situation of the family), or may lead to spillover benefits for other children in the child’s community (e.g., at the same school or in the same neighborhood). Measuring these types of potential spillover benefits is more costly. If these outcomes are not included in the evaluation, it is even more difficult to incorporate them into a cost-benefit calculation.

Second, many of the benefits captured in an evaluation cannot be expressed in monetary terms, either as benefits to program participants or to the rest of society. As illustrated in Table 3.2 (and in the specific studies reviewed in Chapter Four), only a subset of the outcomes that may be affected by an early intervention program can be readily expressed in monetary terms. In other cases, the assumptions needed to assign a monetary value to a given outcome are so heroic that it is preferable to err on the side of undervaluing a program’s benefits. To the extent that cost data are easier to collect and less subject to under- or overestimation, cost-benefit calculations for early intervention programs will likely err on the side of being conservative.

---

9 This assumes that the dollar values assigned to those program benefits that can be monetized are not biased upward or downward.

10 The use of the scorecard still allows the decisionmaker to account for benefits that are not monetized and to use his or her own subjective weights in valuing those outcomes. See Sen (2000) for further discussion of this issue.

11 This conclusion rests on the assumption that cost data are less likely to be underestimated or overestimated. This may be reasonable for those costs directly associated with service provision. However, indirect costs may be equally hard to measure or estimate as some of the benefits listed in Table 3.2. Data collection constraints may also result in underestimation of program costs if not all areas of cost are measured (e.g., cost shifting). However, given that costs are typically incurred during a fixed interval of program provision, while benefits may accumulate indefinitely into the future, the inability to capture the (discounted) monetary value of long-run benefits in certain domains is likely to outweigh the short-term costs that are underestimated. Because this conclusion is by no means universal, whether net benefits are under- or overestimated needs to be assessed on a case-by-case basis.