This chapter applies the methods outlined in the previous chapters to the Starting Early Starting Smart (SESS) program. We consider both data now being collected by SESS and potential options for future data collection and program design. This exercise not only informs SESS policymakers about the use of current data and future opportunities for analysis, but it also helps illustrate how the methods discussed can be put into place for a real-world programs.

We begin this chapter by describing the SESS program. Then we outline approaches to analyzing cost and outcome data for the program. We also discuss some key methodological considerations relevant to conducting cost and outcome analysis for this program.

THE SESS PROGRAM AND EVALUATION DESIGN

SESS is designed to test the effectiveness of integrating behavioral health services for children from birth to age seven and their families, relative to the outcomes for children and families who receive the usual standard of community care. Integrated behavioral health services are defined as substance abuse treatment, substance abuse prevention, and mental health services. The initial four-year phase of the SESS program—Phase I—began in 1997.

---

1This discussion of the SESS program and evaluation design draws on the Starting Early Starting Smart Phase One Report, prepared by the SESS Data Coordinating Center, August 1998.
SESS currently has cooperative agreement grantees in 12 sites nationally. These sites fall into two natural clusters based on their organizational settings—primary health care (PC) and early childhood development (EC). PC sites provide health care to families of target (index) children, and EC sites provide preschool education services to index children. There are currently five PC sites and seven EC sites. (See Appendix A for a full list of SESS sites and a brief description of their program features.) These clusters vary in several important ways, as shown in Table 5.1. PC sites specifically target moderate- to high-risk families. However, participants at EC sites also generally demonstrate relatively high levels of stress and risk factors.

SESS is purposefully designed as a multisite study encompassing diverse field settings in hopes of generating strong evidence of its general applicability. In addition to units of observation at the program level (PC and EC), the units of analysis for the individual level are the index child and the family. The logic behind the design is twofold:

Table 5.1

<table>
<thead>
<tr>
<th>Characteristics of SESS PC and EC Demonstration Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary Care</strong></td>
</tr>
<tr>
<td>Intervention begins from birth to age 3 in most sites</td>
</tr>
<tr>
<td>Eligibility is based on individual screening to target caregivers or children who have specific risk behaviors</td>
</tr>
<tr>
<td>Program focuses behavioral health resources on parent</td>
</tr>
<tr>
<td>Needs of caregiver determine program participation</td>
</tr>
<tr>
<td>Case management component is an innovative addition in this setting</td>
</tr>
<tr>
<td>Experimental design is used for all sites</td>
</tr>
</tbody>
</table>
• Use an experimental or quasiexperimental design to detect program effects at the individual level, and

• Use variation in target population, program context, or program intervention at the program level to explain differences in program effectiveness across sites.

The sample sizes vary across sites, but most are around 100 to 300 index children. The pooled sample consists of 1,584 persons in the treatment group and 1,303 persons in the control (or comparison) group.

The current SESS evaluation is designed to test two specific hypotheses:

• The integration of behavioral health services within PC or EC service sites will lead to higher rates of entry into prevention, early intervention, or the treatment of children/families identified as in need of services (also greater participant satisfaction).

• The integration of behavioral health services within PC or EC service sites will lead to improvements in social, emotional, and cognitive functioning in children and families served.

The first hypothesis focuses on outcomes of services access and utilization and satisfaction, while the second focuses on family functioning, parent-child interaction, and child outcomes.

SAMHSA and CFP have funded a set of cross-site data activities that include data collection, manipulation, and analysis. As part of these activities, they have mandated the creation of an overall program database. The five types of data collected as part of this database include site-level intervention descriptions, contact log data (collected only for the treatment group), Services Access and Utilization and Satisfaction Survey, baseline data, and outcome data. These measures are collected at baseline and for an 18-month follow-up period, with follow-up intervals that average six months (PC sites) or nine months (EC sites). Baseline data and some follow-up data have been collected for treatment and comparison groups. While most sites have attempted to include a comparison group, some sites include no comparison group or a comparison group that receives
some SESS services.\(^2\) CFP and SAMHSA are considering funding a longer-term follow-up for participants in a subset of the current sites. Currently, no cost data are being collected in Phase I, nor are the SESS evaluation design and the longer-term follow-up currently incorporating cost-benefit or related analysis.

CFP and SAMHSA plan to implement a second phase of the SESS program (Phase II), which is currently being designed. Assessing the feasibility of including cost and outcome analysis is part of the planning process for Phase II. In the remainder of this chapter, we assess the utility of data being collected in Phase I for this type of analysis and make recommendations for alterations to the Phase I design, which could be implemented in Phase II.

**USING THE SCORECARD AS A FRAMEWORK**

As a framework for our discussion of potential cost and outcome analyses for the SESS program, we return to the scorecard introduced in Chapter Two of this report (Table 2.1). By characterizing the cells of the scorecard that can be filled in with Phase I data, we can assess the types of analysis that could be conducted with the data currently being collected, and we can identify additional data that would need to be collected in the next phase.\(^3\)

As discussed in Chapter Two, a number of types of cost and outcome analyses could be undertaken for such a program as SESS. Specifically, at least three broad types of analysis could be conducted for this program:

- Cost-savings or cost-benefit analysis, whereby the costs of the program are compared to the benefits of the program from the perspective of the government and society at large, respectively.

- A type of cost-effectiveness analysis, which compares the change effected by different variants of the PC sites or the EC sites or

\(^2\)See Appendix A for more information about each site’s comparison group.

\(^3\)In making our recommendations, we do not explicitly discuss a number of the methodological issues described in Chapter Two, such as choice of discount rate and accounting for statistical and scenario uncertainty. These can be addressed during the cost analysis, following standards established in the cost-benefit literature.
Cost and Outcome Analysis of Starting Early Starting Smart

examines which design features of SESS programs were associated with the greatest “bang for the buck.”

- Characterization of the costs of implementing SESS so that future sites hoping to replicate the program have reasonable expectations regarding the costs they would incur.

While other approaches could certainly be enumerated, these three represent the general classes of analysis best aligned with the stated objectives of the policymakers for this program.4

As we proceed in the remainder of this chapter, we rely on the scorecard framework to make a series of recommendations about the evaluation design and the collection and analysis of cost and outcome data. However, a number of our recommendations specific to cost and outcome data depend, in part, on the type of analysis desired for the SESS program. This in turn will reflect the objective that the analysis is trying to achieve, such as the three listed above. For example, if the goal of the cost and outcome analysis is to characterize the costs of implementing SESS for potential future replication, the bulk of the cost data would pertain to the costs to the agency implementing the program. However, if the goal is a comparison of the costs and benefits of the program from the perspective of society at large, then a more comprehensive enumeration of the costs and outcomes of the program would be required. We revisit these issues again at the end of this chapter.

**Recommendation**: Specify the explicit goals of the cost and outcomes analysis to guide the scope of cost and benefit data collection and analysis.

**Defining the Baseline and Alternative Policies**

We first need to establish the columns of the scorecard—i.e., what would serve as the baseline comparison group and what would serve as the alternative programs. As discussed above, the baseline repre-

---

4These objectives and other issues related to the application of cost and outcome analysis to SESS are described in Cannon, Karoly, and Kilburn (2000). This document summarizes a meeting held between SESS funders and program staff and experts in cost analysis from both RAND and other organizations.
sents the world without the SESS program elements. In the case of the SESS Phase I design, there is a baseline case associated with the two basic program models: primary care (PC) sites without SESS and early childhood (EC) sites without SESS. An SESS information packet states that grantees are required “to address the multiple needs of poor and at-risk families and their very young children by providing coordinated, wraparound services, with special emphasis on services that address the participants’ behavioral health needs.” Hence, the marginal contribution of SESS is the integrated mental health and substance abuse prevention and treatment services delivered in these settings, plus coordination activities that may change the amounts of other services that participants receive. SESS’s marginal contribution is not the entire range of services provided at these sites. This is why the comparison group is PC sites or EC sites without SESS rather than a control group that receives no services of any type, including PC or EC services.

The alternative programs under consideration are the PC and EC sites with SESS. However, the Phase I demonstration of SESS was purposefully designed to have variation within the PC and EC models in the treatment populations and suite of services offered to participants across the demonstration sites. As a result, there is a baseline for each combination of geographic site and program model. Thus, it would be possible to consider a number of variations of SESS PC and EC sites to assess how differences in the population served and/or the services provided influenced costs and outcomes. This corresponds to the second type of cost and outcome analysis—cost-effectiveness analysis—enumerated above.

For the sake of brevity, in the remainder of this discussion we will assume that for our hypothetical example there is only one variant of

---

5 It is also possible to design an evaluation with a baseline that represents a world with no program at all, either the basic services offered at PC or EC sites or any of the add-on elements of the SESS program. In this case, the costs and benefits of both basic PC or EC services plus the SESS overlay would be compared with a control group that received no SESS, EC, or PC services.

6 In the Phase I implementation of SESS, those in the control or comparison group at the PC sites receive services from the same PC provider that also offers integrated SESS services to the treatment group. It is possible that even the basic PC services are changed as a result of the provider offering the integrated SESS services for the treatment group, for example, stemming from the capacity building of the staff, and so on.
PC with SESS (PC plus SESS) located in one geographic site, but two variants of EC with SESS in two separate geographic sites, which we shall call EC1 plus SESS1 and EC2 plus SESS2. In Table 5.2, we show how the columns in the scorecard would appear for this set of comparisons.

The consideration of comparison groups and policy alternatives raises four design issues for the planned Phase II evaluation of SESS. The first is the use of an experimental versus quasiexperimental design, i.e., whether the baseline is a randomly assigned control group or a matched comparison group. The Phase I design (see Table 5.1) includes a mix of sites, some with random assignment (primarily PC sites) and others with matched comparison groups (mostly EC sites). Preliminary data from the evaluation raise concerns about the preintervention comparability of the matched comparison groups in the EC sites (see the summary of the discussion in Cannon, Karoly, and Kilburn, 2000). If such differences exist, any postintervention differences between the treatment and comparison groups may be due to other factors besides the SESS services. To obtain the best research results, random assignment would be used for the evaluation design at all sites in a subsequent demonstration phase, if at all possible.

However, random assignment may not be feasible for several reasons. As we pointed out earlier, results of early childhood interventions can be extremely sensitive to the risk characteristics of the population they serve. They may have big effects when applied to high-risk children, but smaller effects when applied to lower-risk

\begin{table}[h]
\centering
\caption{Illustrative Scorecard for Hypothetical SESS Example: Alternative Policies}
\begin{tabular}{|l|c|c|c|c|c|}
\hline
 & PC Only & PC+SESS & EC1 Only & EC1+SESS1 & EC2 Only & EC2+SESS2 \\
\hline
Impacts & & & & & & \\
\hline
Program Descriptors & & & & & & \\
Cost Elements & & & & & & \\
Outcomes & & & & & & \\
\hline
\end{tabular}
\end{table}
Assessing Costs and Benefits of Early Childhood Interventions

Random assignment means refusing program services to some high-risk children, and this may be difficult to do in certain settings. In the case of the SESS program, this may be more of a concern for the PC sites where treatment and control children are served by the same provider. At EC sites, this may be less of a concern, since the SESS services are offered to whole classrooms of children rather than to randomly selected individuals. Likewise, control groups consist of whole classrooms to which SESS services are not offered. One can reasonably expect to find children at all risk levels in both the control and participant classrooms. But at both kinds of sites, if random assignment is not possible, it is important to match controls to participants in terms of risk factors.

**Recommendation**: Where possible, use random assignment to define control groups in order to provide a more valid test of SESS program effects. When random assignment is not possible, strive to match children in the treatment and comparison groups in terms of their risk factors.

A second issue concerns data collection for the control group. In Phase I, participants and controls alike received an initial interview and several follow-up interviews at intervals that average six to nine months for PC and EC sites, respectively. For each participant, however, each SESS site keeps a contact log that describes every telephone contact and every face-to-face contact with SESS staff. Data this complete and detailed are not available for controls. In particular, it is not known, save by self-report after delays of several months, just what services the controls are receiving. They may, in fact, be receiving many of the same services as the participants. It might be possible to obtain more complete and accurate records of services received by controls from records kept by the service providers. Of course, controls would have to provide consents for SESS to gain access to these records.

**Recommendation**: Strive to collect service, cost, and outcome data on the control groups that are as complete as the data on the treatment groups.

A third issue concerns the extent of variation in the SESS program models as implemented across demonstration sites, both in terms of the services provided and the target population served. In the Phase
I design, the program models, and to some extent the population served, vary by geographic site even within the PC and EC program models. This variation can be useful for identifying the most successful program designs based on the Phase I outcomes data. However, it is difficult to disentangle differences in program effectiveness stemming from the program model, geographic site, or population served. For Phase II, there are advantages to considering a more limited set of the best designs that emerge from Phase I, possibly implementing the same program model in two geographic sites or for different target populations or implementing two different models in the same geographic site or for the same target population. Alternatively, it may be desirable to fix the target population, selecting among the at-risk groups identified in Phase I that benefit the most from the SESS program model. In either case, for an evaluation of a given total sample size, a more refined and uniform program model in Phase II will allow the evaluation to consider how outcomes and costs vary with the characteristics of the site, target population served, or program model. This will be important information to guide future program implementation.

**Recommendation**: In Phase II, impose more uniformity in the program models across sites, strategically selecting a few variations in design based on outcomes data from Phase I.

A fourth important consideration that influences the viability of conducting cost and outcome analysis for SESS is the ability of the Phase I or planned Phase II evaluations to retain subjects (both control and treatment group members) across time. This is important because attrition from evaluation studies is rarely random. Instead, those who continue to receive program services or to be assessed in terms of their outcomes are likely to differ from those who drop out of the program or are lost to follow-up in ways that may not be controlled for by differences observable to the researcher. Analyzing data that contain only the individuals who remain in the program over time and who continue to be monitored could generate misleading conclusions regarding the effectiveness of the program. In the first follow-up of Phase I data collection, participant retention from the initial survey ranged from nearly 99 percent to a low of 56 percent across sites, with mean retention in the EC sites and PC sites of 82
and 61 percent, respectively. Because of the importance of collecting long-term outcomes for children’s intervention programs, this issue also merits special attention during the Phase II design.

**Recommendation**: Use the information from the Phase I evaluation to assess the reasons for attrition from the study. In Phase II, devote more resources to retaining study subjects, remedying the retention problems identified for some sites in Phase I.

**Describing SESS Sites**

Now we turn to filling in the rows that should be described under the three broad headings in Table 5.2. The first information we need to specify are the features, or “program descriptors,” of each baseline program and each alternative policy. They should be detailed enough so that future sites, which may be considering implementing variations of the policies, could have a reasonable expectation of replicating the conditions under which the costs and outcomes were realized.

While a complete list of program descriptors may include dozens of entries or more, we list types of information here that would be candidates for inclusion:

- Population served, especially including risk category or characteristics that determine risk. Eligibility criteria should be listed as well.
- Characteristics of personnel providing services (such as education, certification, and bilingual skills).
- Typical services received by participants (such as a particular substance abuse prevention curriculum, enriched preschool that focused on specific skills, psychiatric evaluation, medication monitoring, and residential substance abuse treatment).
- “Dosage” of services, including number of visits and length of visits of various types. Note that services provided will generally

---

7 Documentation provided by the SESS Data Coordinating Center based on response rates as of December 12, 2000.
be tailored to the population served, so types and dosages of services will need to be specified separately for different population subsets.

We indicate some illustrative program descriptors for our hypothetical SESS example in Table 5.3. Note that ideally, the features of the “baseline” or comparison program should be as close as possible to those of the “treatment” program, save for the specific features that characterize the SESS program.

When characterizing the program features, it is important that they be based on information on how a program is actually implemented, not just on the planned design. In the Phase I evaluation of SESS, a component of the data collection includes site visits to gather information about how each program model is actually operating. This is critical information required for conducting a valid comparison across program models and should be continued in the Phase II

Table 5.3

Illustrative Scorecard for Hypothetical SESS Example:
Program Descriptors

<table>
<thead>
<tr>
<th>Program Descriptors</th>
<th>Impacts</th>
<th>Alternative Policies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population served</td>
<td></td>
<td>PC Only PC+ SESS</td>
</tr>
<tr>
<td>Age of child at enrollment in program</td>
<td></td>
<td>EC1 Only SESS1</td>
</tr>
<tr>
<td>Eligibility criteria</td>
<td></td>
<td>EC1+ Only SESS2</td>
</tr>
<tr>
<td>Transportation provisions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child care available during parent appointments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health services provided</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Etc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost Elements</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcomes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


design. This information is also useful for ensuring fidelity to a program model as designed, so that “program drift” is minimized and “dosage” levels are maintained.

Recommendation: In Phase II, continue to collect information on program features through site visits and other mechanisms to characterize accurately features of the intervention models as they are implemented and to ensure fidelity to the program model.

COLLECTING AND ANALYZING SESS PROGRAM COSTS

The second broad heading shown in Table 5.2 is cost elements. The cost of the SESS program would entail a comparison, for each program model, of the costs with and without the SESS component. That is, the costs of the PC plus SESS programs would be the difference between the costs of the PC model without SESS and the costs of the primary care model with SESS. Similarly, the costs of the EC plus SESS programs would be the difference between the costs of those programs with and without the SESS component. This comparison thus requires collecting cost data for both treatment and control group participants at each site where SESS is implemented. Collecting cost detail at the level of each participant is possible, but this can be time-consuming. It is probably sufficient, for most analyses of SESS that would be of interest, to construct aggregate program costs at each site, rather than cost disaggregated by participants or groups of participants at each site. The most likely exception would be if high-risk children were provided much more intensive services, or were retained longer in the program, than low-risk children.

Recommendation: Collect cost information for both treatment and control groups at each site where SESS is implemented.

The cost principles outlined in Chapters Two and Three should guide the completion of this section of the scorecard. In particular, information characterizing the following categories should be enumerated in the scorecard:

- Resource categories. These include personnel, equipment, facilities, and supplies/other.
- Explicit expenses and in-kind costs.
Cost and Outcome Analysis of Starting Early Starting Smart

- Fixed and variable costs.
- Consumable and nonconsumable items.
- Investment costs and operating costs.
- Stakeholder group. Such as participants, the agency implementing the program, or society at large.

Rather than including a row for each combination of these various categories, a good start would be to include sections for stakeholders and resource categories. As discussed earlier, the following groups are likely to incur costs as a result of the program:

- Participants. Their costs may include time and resources getting to appointments, child care while the parents are in meetings or appointments, the value of the time spent in appointments, and others.

- The Agency Implementing the Program. The agencies’ costs will include the labor bill for staff, the rent or space costs, such operating costs as utilities, supplies and equipment, and others.

- Other Agencies or Providers. These may include public or private agencies or providers to whom SESS participants are referred for services, such as special education services or family violence prevention programs.

- Society as a Whole. The costs to other components of society might be the value of the time of volunteers at the agency implementing SESS, donated space or supplies, or the value of the public infrastructure, such as public transportation, which may play a role in the delivery of SESS services.

We have shown these four groups of stakeholders, which might accrue costs, in italics in Table 5.4. As noted in Chapter Three, it is critical that identical cost information be collected for both treatment and control groups for each of the parties listed above. This allows investigation of possible cost-shifting or cost-offsets that otherwise might go undetected.

Also in this table, we have listed a few examples of resource categories for the two groups of stakeholders—participants and agencies implementing the program—as an illustration. We have also
included a couple of examples of specific items, which might be included in the rows. Since participants are unlikely to incur facilities costs or equipment costs as a result of participating in SESS, we have only included personnel and supplies/other categories for participants. A much richer list of cost entries would need to be developed for each stakeholder and each resource category as part of the analysis of the SESS program. Once the particular items that go in the rows have been identified, they can be demarcated according to the other characterizations enumerated above, such as explicit expense or in-kind expense, investment cost versus operating cost, and so on.

**Recommendation**: The cost information should be as comprehensive as possible. Costs borne by various parties by resource category should be differentiated; the time period that costs are incurred should be identified; and direct and indirect costs, fixed and variable costs, and goods and services provided in-kind should be measured.

Currently, SESS data collection efforts in Phase I focus on outcome measurement and do not include data on costs. Even though such issues as the quality of comparison groups are not likely to be resolved in Phase I, collecting cost information for the extension sites in Phase I would still have great utility, particularly for informing the Phase II design. For instance, if different types of PC plus SESS or EC plus SESS sites realized similar outcomes, but one type of either PC or EC site had half the costs of the others, policymakers may want to focus Phase II investments in the lower-cost option. Similarly, collecting data in the Phase I extension sites might help identify specific program features that have the greatest impact on key outcomes in relation to cost per family served. Again, this could help suggest which program features Phase II should emphasize or encourage. Beginning to collect cost data for the Phase I extension sites would have the additional advantage of serving to work out data collection procedures before Phase II, and to indicate how much of the Phase II evaluation budget should be set aside for the collection and analysis of cost data.

**Recommendation**: Collect cost data for the Phase I extension sites to inform the design of Phase II and help prepare for Phase II cost data collection.
### Table 5.4

*Illustrative Scorecard for Hypothetical SESS Example: Cost Elements*

<table>
<thead>
<tr>
<th>Alternative Policies</th>
<th>PC Only</th>
<th>PC+ SESS Only</th>
<th>EC1 Only</th>
<th>EC1+ SESS1 Only</th>
<th>EC2 Only</th>
<th>EC2+ SESS2 Only</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Program Descriptors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cost Elements</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personnel costs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lost work time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supplies/other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transportation resources</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child care costs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agency Implementing Program</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personnel costs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of hours spent per nurse home visit, including preparation, travel, follow-up, etc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of nurse home visits per participant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equipment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer and related equipment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facilities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utilities used</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supplies/other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Travel costs for nurse home visit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other agencies or providers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Society as a whole</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Finally, in collecting cost information, whether for Phase I extension sites or Phase II sites, it is important that the data collection procedures be as uniform as possible across SESS demonstration sites, with all sites capturing costs for the same parties, cost elements, and time periods. This is implicit in the construction of the scorecard, yet it is still worth emphasizing given that the capacity for data collection
Assessing Costs and Benefits of Early Childhood Interventions

and the cost accounting systems may be quite different across sites. A critical element in the collection of cost data will be appropriate training and support at each site and for any data collection organization that may operate across sites. The cost associated with training for and gathering cost information (and the outcome information discussed below) should also be collected. If data collection becomes a standard part of implementing the SESS model, this information will allow these costs to be incorporated into the estimate of the full program costs. Alternatively, if future implementation of SESS will not require detailed data collection, or only a more streamlined data collection procedure, the program costs can be adjusted accordingly. The same is true for the cost associated with the analysis of the cost and outcome data collected.

**Recommendation**: Plan for proper training and technical support of SESS sites and any cross-site data collection organizations to ensure uniformity in the collection of cost data. Collect information on the cost of data collection, training and support, and the related analyses of the data.

COLLECTING AND ANALYZING SESS PROGRAM BENEFITS

The final heading shown in Table 5.2 is program outcomes. Like the cost elements in the scorecard, the outcomes in the scorecard would also need to demarcate the individuals to whom benefits accrue and the period when gains are realized. The benefits of early childhood intervention programs have typically been measured for program participants in the four broad domains reviewed in Chapter Three: emotional and cognitive development, education, economic well-being (e.g., public assistance receipt, income, crime), and health. The specific outcome measures in each category—and whether they are measured for participating children, parents, or both—is a function of the program design and the expected areas of impact. As noted in the discussion in Chapter Three, some of these impacts—such as those in the economic sphere and a subset of those in the education domain—when applied to children require longer-term follow-up to observe changes in their outcomes at more advanced ages, long after the intervention has ended.
Chapter Three also highlighted some impacts that result from changes in participants’ behaviors that can also affect outcomes for nonparticipants. For example, reduced criminal activity on the part of participating parents or children produces benefits to other members of society in the form of lower crime rates. Another example: Improved behavior of program participants during their school-age years may improve classroom learning for other children at school. Likewise, improved outcomes for the parent may have spillover benefits for the parent’s other children in addition to the target child in the intervention.

The current data collection effort for the first phase of SESS is guided by the expected areas of program impact and an evaluation initially planned based on a two-year period of data collection. In particular, the SESS evaluation focuses on multiple domains of expected impact: access, utilization, and satisfaction with behavioral health services and family functioning; parent-child interactions; and child outcomes. Data currently being collected include measures of the following:

- Focal child characteristics.
- Family/household characteristics.
- Parent/caregiver characteristics, such as demographics, education, employment, public assistance, insurance, etc.
- Child problem behavior and social skills.
- Child cognitive development.
- Parent-child interaction.
- Parent/caregiver stress and negative/positive behaviors.
- Parent/caregiver mental health problems.
- Home environment, such as safety/violence and learning opportunities.
- Service utilization and satisfaction.

As indicated by this list, the SESS evaluators are collecting outcome data for both parents and children.
For purposes of the various cost and outcome analyses, the outcomes being collected for the SESS evaluation do not include most of the measures italicized in Table 3.2, i.e., those most readily translated into monetary benefits, either to government (taxpayers) or to other members of society. In fact, many of the above outcomes—which largely fall in the class of cognitive or emotional development measures—would be difficult to translate into monetary terms. Other benefits, such as better access to needed services or more appropriate use of health care services, are also difficult to express in monetary terms. This makes a formal cost-benefit or cost-savings analysis problematic in that only a limited set of outcomes might possibly be valued in dollar terms to be compared with program costs. Unless the program impact for those outcomes that are monetized is very large and favorable, so that sizable dollar benefits are generated, it is unlikely that a cost-benefit analysis would show a favorable outcome for the SESS program based on the information available after two years.

Given the current data collection plan for Phase I, cost-effectiveness analysis, which compares the change in outcomes elicited by a program to the costs of the program, is feasible provided cost data are assembled for the current or extension sites. This is because the outcomes are not translated into dollar terms but rather remain in their natural units, such as values on a given scale. Because no summary cost-benefit measure is generated, however, this approach requires decisionmakers to weight the various outcomes using their own subjective weights. Another type of analysis, which could be executed with the currently available data, is an assessment of which design features of programs yielded the greatest influence on outcomes. This type of analysis is currently planned as part of the Phase I evaluation.

**Recommendation**: If Phase I cost information can be collected as recommended above for Phase I extension sites, focus cost and outcome analysis based on Phase I data on cost-effectiveness measurement.

If the objective of the cost and outcome analysis is to perform cost-savings or cost-benefit analysis, it will be important to broaden the types of short-term measures collected, especially for parents and other caregivers, and to consider an evaluation with a longer-term
follow-up. As demonstrated by the cost-benefit analyses in Chapter Four, parents’ outcomes have the potential to produce the largest short-term gains as the result of an early childhood intervention program. In contrast, improvements in children’s outcomes may take years or even decades to reveal themselves. For this reason, if analysis that compares the benefits and costs of SESS is desired, collecting longer-term outcomes in Phase II would be valuable. While modeling is able to predict some longer-term outcomes based on observed changes in outcomes in the short run, obtaining data over the longest period possible avoids the statistical uncertainty inherent in such forecast modeling. The scenario uncertainty remains, of course.

A possible longer-term follow-up of the Phase I or planned Phase II demonstration sites would allow for a broader set of measures to be collected for participating children and their parents, including those that might produce larger impacts or impacts that can at least be monetized. The cost-benefit analyses of the early childhood programs reviewed in Chapter Four demonstrate the value of collecting information in the short- and medium-term (e.g., two to 10 years) for parents and in the longer-term (e.g., 10 to 20 years) for children on outcomes such as public assistance program use, employment, earnings, and criminal activity. If behavioral changes are large in these areas as a result of the SESS intervention, they can produce sizable dollar benefits that, even when discounted, will be a large offset to the costs of the program.

Table 5.5 illustrates some of the outcome measures that might be used for longer-term follow-up of the SESS program. The key outcome areas discussed in Chapter Three that are easily expressed in dollar terms are represented, and measures for both children and adults are assessed as of a specific age, $A$, of the focal child. Whether or not the SESS program will produce outcome gains in these areas has yet to be determined, but there is reason to believe that increasing access to substance abuse treatment services and mental health services will affect at least some of these domains. Substance abuse has been found to impose huge economic costs on society (Rice et al., 1990), and treatment has been demonstrated to be more effective than either no treatment or incarceration (McLellan et al., 1996). Other research has found that over 90 percent of the total cost savings produced by substance abuse treatment is in the form of
reduced criminal justice system costs (see, e.g., CSAT, 1999). Moreover, in a comparison of treatment to other cocaine control programs, Caulkins et al. (1999) showed that treatment was more cost-

<table>
<thead>
<tr>
<th>Table 5.5</th>
<th>Illustrative Scorecard for Hypothetical SESS Example: Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Impacts</strong></td>
<td>PC</td>
</tr>
<tr>
<td>Program descriptors</td>
<td>Only</td>
</tr>
<tr>
<td>Cost elements</td>
<td>Outcomes for children</td>
</tr>
<tr>
<td></td>
<td>Number of grades</td>
</tr>
<tr>
<td></td>
<td>repeated through age A</td>
</tr>
<tr>
<td></td>
<td>Months employed through age A</td>
</tr>
<tr>
<td></td>
<td>Months receiving public assistance through age A</td>
</tr>
<tr>
<td></td>
<td>Emergency room visits through age A</td>
</tr>
<tr>
<td></td>
<td>Etc.</td>
</tr>
</tbody>
</table>
effective than other approaches, including prevention, enforcement, and interdiction.

Improvements in mental illness rates would be expected to yield gains in labor force outcomes given that the percentage of persons out of the labor force and unemployment rates are significantly higher for persons with mental disorders (Sturm et al., 1999). The most comprehensive evidence on mental health services that explicitly incorporates cost-outcome methods is for the assertive community treatment (ACT) program, which provides services for those with serious mental disorders. Results indicated that subjects in the experimental group had improved outcomes compared to the control group and that family and community burden did not increase. Given increased wages and lower income support for the experimental group, societal costs were found to be slightly lower than for the control group (Test and Stein, 1980; Stein and Test, 1980).

These findings were countered by results that showed the results of the two groups converged after the program was terminated (see discussion in Hargreaves et al., 1998). Given the focus of the SESS intervention on increased access to and utilization of substance abuse and mental health treatment services, the SESS program could also produce benefits in similar areas.

It may also be fruitful to collect information in other outcome domains for possible inclusion in a cost-benefit analysis. For example, information on educational outcomes for children may be collected as early as the primary grades, with possible improvements in such outcomes as lost school days, grade repetition, and special education use that can be valued and tallied against program costs.

For parents and other caregivers, improvements in physical and mental health or reductions in such outcomes as family violence and child abuse and neglect may be evident in the short and medium term. These outcomes can potentially be valued as well in terms of increased labor market productivity and reduced use of other health care services. Again, it is not certain that the SESS program will significantly affect these outcomes, but they are among the likely candidates for improvement, and they can be translated into monetary benefits for participants or other members of society. Given the opportunity costs associated with added data collection, any new
measures collected should be selected based on a theoretical model of the SESS program’s expected impacts along with evidence that similar interventions have produced gains in those areas.

**Recommendation**: If cost-benefit or cost-savings analysis is the objective for SESS, then outcome data should be supplemented to include information for parents and other caregivers in the short and medium term in the domains of health and economic well-being (e.g., labor market outcomes, public assistance use, criminal activity, and justice system contact) and for children in the medium term in the domain of educational outcomes and longer term also in the domain of economic well-being. The choice of specific outcome measures should be guided by findings from related evaluation studies whenever possible.

If a longer-term evaluation study is designed or anticipated for either Phase I or Phase II, several methodological issues discussed in Chapter Three should be considered. First, if a long-term follow-up is anticipated at the outset of the evaluation, it is important to collect information that will ensure the lowest possible rates of attrition and that allow data collection through administrative sources along with survey data. This would include, for example, obtaining identifying information for program participants, such as Social Security number or driver’s license number, at the outset of the intervention. This would allow tracking of those in the treatment and control groups for subsequent follow-up interviews or searches for data in administrative databases (e.g., employment histories, criminal records).

**Recommendation**: For a Phase I follow-up or Phase II design, obtain information from participants that allows collection of administrative data and permits effective tracking of individuals to increase response rates at later follow-ups.

Second, as discussed in Chapter Three, it is desirable to collect complete histories for some outcomes that may generate a continuous flow of dollar benefits. Thus, for example, if employment outcomes are better each year after an intervention ends, it would be ideal to know about employment rates in each year since the last follow-up in addition to their current status. A complete history of public assistance program use or use of costly special services in education or health care would also be relevant. Depending on the interval since
the last follow-up, it may be difficult for respondents to recall a complete history, but such retrospective information can be of high quality when the events recorded are particularly salient. Administrative data, when available, also often provide a complete history with less concern about possible recall bias.

**Recommendation**: Where possible, collect complete histories using retrospective survey questions or administrative data for outcomes that may generate a continuous flow of dollar benefits (e.g., labor market outcomes, public assistance program use, use of costly health or education services).

Third, it may be possible—for some outcomes affected by the SESS intervention—to forecast future benefits beyond the period of follow-up. For example, the cost-benefit studies reviewed in Chapter Four projected future earnings beyond the last follow-up based on the earnings histories of participants observed to date. This allows estimates of increased tax revenue to be projected beyond the last period that participants’ outcomes are observed. Likewise, the reduction in future criminal activity and welfare program use was forecast based on observed behavior as of the final follow-up. In other areas, such forecasts may be possible although the methods to do so may require further development. For example, it may be possible to model the link between children’s early cognitive gains (e.g., in IQ or achievement tests) and their economic success as adults. We are not aware of any cost-benefit studies that have made such a projection but it should be feasible given other sources of data that would permit estimation of this relationship (see, for example, Currie and Thomas, 1999).

**Recommendation**: When supported by other empirical evidence, project future benefits based on observed outcomes. Consider additional method development that would permit such forecasts for a broader range of outcomes.

**COMPARING COSTS AND BENEFITS OF SESS**

The preceding discussion has made it clear that the choice of what type of cost and outcome analysis will be conducted is a driver of the data collection and issues that need to be addressed in preparation
of the scorecard. Thus, as indicated by the first recommendation in this chapter, it is important to specify the explicit goals of the cost and outcome analysis in order to determine the nature of the cost and outcome data required. We now briefly summarize the feasibility of undertaking each of the three options outlined at the beginning of this section, given current data collection efforts, and describe some of the changes to data collection that would be required to undertake each of the options in Phase II.

**Cost-Benefit or Cost-Savings Analysis**

This is the analysis option that would require the greatest modifications to the current data collection plan. This is primarily because under cost-benefit or cost-savings analysis, the analyst would attempt to convert benefits to a monetary value to compare with costs, and the outcomes currently being measured do not lend themselves well to being expressed in monetary value. Hence, to undertake this type of analysis, the types of outcomes collected would need to be expanded as would the duration of the follow-up. Needless to say, cost data would also need to be collected.

This approach would not only take the longest amount of calendar time to execute, as analysis could only get under way after some follow-up time elapsed, but it would also be likely to require the largest budget of the analysis options. This is because new outcomes measures would need to be developed along with a data collection plan for costs. A plan for minimizing participant attrition would need to be devised as well.

This is likely to be the best analysis option only if program sponsors are committed to answering the unique questions addressed by this approach: whether SESS benefits “pay” for their costs, either from the perspective of the government or society as a whole. If this analysis is pursued, it is also important to recognize that the monetary estimates of program benefits are likely to be conservative. Consequently, the program impacts in those domains that can be monetized must be sufficiently large, and sustained over a long enough period, to generate benefits that exceed program costs. The conservative nature of the benefit calculations may produce disappointing results, especially when only short-term results are available. The program may only appear to be cost-beneficial when the evaluation
has incorporated information about program outcomes observed a decade or more after the intervention has ended.

**Cost-Effectiveness Analysis**

Cost-effectiveness analysis for SESS would primarily entail supplementing current data collection with cost data. Not as formidable as the changes required to implement cost-benefit analysis, collecting cost data nevertheless entails large time and resource investments in either or both of the Phase I extension sites and Phase II sites.

This option would answer questions about the relative effectiveness of implementing SESS at PC or EC sites, whether targeting the program to particular participants made a difference, and which treatment components yielded the greatest gains. All of these could be compared on a per-dollar basis if accompanied by cost data.

**Replication Analysis**

The final type of cost and outcome analysis, which could be undertaken for SESS, is an assessment of the cost of implementing the program in additional sites. This would be most valuable if policymakers envisioned ”scaling-up” SESS in the future or if they expected that other agencies might begin to implement the program. If future expansion of the program to other sites is not anticipated, this option has little merit.

This analysis would require collecting cost data, as in the other two analysis options. However, unlike in the cost-benefit or cost-effectiveness options, it would not be particularly important to collect outcome data. It would be important to include program descriptor information, because this would help future sites gauge the comparability of their setting to SESS demonstration sites.

In sum, there is no right or wrong answer to the type of cost and outcome analysis undertaken for SESS. The objectives of the consumers of the analysis dictate the approach taken, which in turn has implications for the collection and analysis of data on program costs and benefits. Clearly, program decisionmakers may have to make trade-offs between what they might like to achieve and how much of a resource commitment they are willing or able to make.