EVALUATING PROGRAMS FOR VULNERABLE CHILDREN AND YOUTH

by

Adele Harrell

LASHC Paper Series No. 3

July 1996

Papers prepared in this series are not formal publications of the World Bank. They present preliminary and unpolished results of country analysis or research that is circulated to encourage discussion and comment; any citation and the use of this paper should take account of its provisional character. The findings, interpretations, and conclusions expressed in this paper are entirely those of the author(s) and should not be attributed in any manner to the World Bank, its affiliated organization, members of its Board of Executive Directors or the countries they represent.
TABLE OF CONTENTS

Introduction ............................................ 1
Clarifying the Evaluation Questions ......................... 4
Developing a Logic Model .................................. 5
Reviewing Readiness for Evaluation .......................... 10
Selecting an Evaluation Design ................................ 11
   Experimental Designs .................................. 11
   Quasi-Experimental Designs ............................. 17
   Non-Experimental Evaluations ......................... 23
Performance Monitoring .................................... 28
Case Study/Ethnography ................................... 33
Cost Studies ........................................... 38
Choosing among the Impact Designs ......................... 45
Identifying Potential Evaluation Problems ..................... 47
Conclusions and Recommendations ......................... 51
Exhibit A. Logic Model Used in Evaluation of the Children at Risk Program 53
Exhibit B. Process for Selecting Impact Evaluation Designs .......... 54
Introduction

Evaluating programs for vulnerable children and youth is both important and challenging. Evaluation is important because governments and non-profit organizations are increasingly recognizing the potential benefits of programs designed to develop human capacity by providing services earlier, rather than later, in the life cycle. Policy makers and service providers want to know what programs accomplish, what they cost, and how they should be operated. They want to know which programs work to combat specific problems and meet the needs of specific population groups or communities. And, increasingly, they want to base these decisions on hard data and evidence, rather than testimonials and impassioned pleas. Evaluating these programs is challenging because the programs address a wide diversity of problems and possible solutions and often include multiple agencies and clients, changing over time to meet shifting service needs.

Vulnerable children and youth include youngsters from birth through late adolescence, with widely varying developmental needs and capacities and widely varying social supports available to foster healthy development. By vulnerable, we mean those who face elevated, identifiable risks of failing to develop into healthy, economically self-sufficient adults. A developmental perspective sees problems in adulthood as the end result of a cumulative process which begins in early childhood. Underlying many programs for vulnerable children and youth is the proposition that early identification of risk and intervention using strategies that are appropriate for the age and level of risk increase the chances that those served will be able to overcome the risks and become productive citizens.

One framework developed for use in designing and evaluating programs to help youth avoid negative outcomes proposes four components of a risk definition: risk antecedents, risk markers, problem behaviors and outcomes.\(^1\) Risk

---

antecedents are those environmental conditions, such as poverty and family
dysfunction, that consistently predict subsequent negative outcomes for young
adolescents. Risk markers are behaviors or conditions associated with more
serious outcomes and which the research literature indicates may be signals of
impending dysfunction. Risk outcomes are defined as those serious, negative
consequences that are most likely to follow from a combination of risk antecedents
and problem behaviors, and are signaled by risk markers. These outcomes include
pregnancy, too-early parenthood, poor pregnancy outcomes, sexually transmitted
diseases, homelessness, prostitution, substance abuse, dropping out of school,
criminal behavior, depression, suicidal thoughts and behavior, death or injury,
accidents and physical and sexual abuse. These outcomes generally result from
problem behaviors such as early sexual activity, truancy, running away from home,
early use of illegal substances and associating with delinquent peers.

For program development and evaluation, risk factors can also be grouped
by the location of the risk (and thus target of the intervention) as: 1) environmental
factors, such as poverty, over-crowded housing, high crime rates, and transient
neighborhoods; 2) family risk factors, such as a family history of alcoholism, crime
and drugs and poor or inconsistent parenting; 3) peer group risk factors, such as
gang membership and problem behaviors among friends; and 4) individual risk
factors, including academic failure and poor attachment to school, early

Risk reduction programs may involve: 1) primary prevention programs
designed to avoid negative outcomes in populations which have, at least
statistically, a high probability of experiencing problems (risk antecedents); 2)
secondary prevention programs to provide help and alternatives for youth with
early evidence of the onset of problems (risk markers); and 3) tertiary prevention
programs that provide treatment and remediation for those experiencing problem behaviors or negative outcomes. The range of service providers and funders is broad and may include a number of different government agencies, private and nonprofit organizations, local community groups or a combination of these organizations. The services may be geographically concentrated in a single area or provided nationwide. The target population may be defined narrowly or broadly in terms of level of risk, developmental stage, and eligibility requirements. The common thread in these undertakings is the shared goal of maximizing the potential of children and youth through a program of services.

This paper reviews the process of developing evaluation designs for these programs. It signals common pitfalls, identifies constraints that need to be considered, and presents ideas for solving some potential problems. Topics and examples were chosen to alert planners to potential design considerations, not to provide a comprehensive review of data collection methods or analysis. Some technical problems and solutions are included by reference; others will require more detailed examination after design decisions have been made. The first step in the process of selecting evaluation designs is to clarify the evaluation questions.
Clarifying the Evaluation Questions

The design of any evaluation begins by defining the audience for the evaluation findings, what they need to know, and when. These questions generally will determine what kinds of evaluation activities should be included in the design.

The major types of evaluations are:

Impact evaluations focus on questions of causality -- did the program have its intended effects? If so, who was helped and what activities or characteristics of the program created the impact? Did the program have any unintended consequences, positive or negative?

Performance monitoring provides information on key aspects of how a system or program is operating and the extent to which pre-specified program objectives are being attained (e.g., numbers of youth served compared to target goals, reductions in school dropouts compared to target goals). Results are used by service providers, funders, and policy makers to assess the programs performance and accomplishments.

Process evaluations are undertaken to answer questions about how the program operates and document the procedures and activities undertaken in service delivery. Such evaluations are extremely useful in identifying problems faced in delivering services and strategies for overcoming these problems.

A comprehensive evaluation will include all of these activities. Sometimes, however, the questions raised, the target audience for findings, or the available resources dictate focusing on only one or two of these activities. Each type of evaluation can include questions related to program costs.

Evaluations also vary in the extent to which preliminary findings are used in program development and change. Formative evaluations are those in which findings from impact evaluation, performance monitoring, and process evaluation are made available to program staff for use in improving program operation and developing additional services. Preliminary results can be effectively used to identify operational problems and develop the capacity of program staff to
conduct on-going self-evaluation and monitoring activities. This use of evaluation findings presents a challenge to evaluators who are faced with estimating the impact of an evolving intervention, which requires on-going measurement of the types and level of service provided. However, this disadvantage is often offset by the gains to the program and the increased utility of the evaluation.

During or immediately after clarifying the evaluation questions, it is important to develop a logic model that lays out the expected causal linkages between the program (or program components) and the program goals. At the same time, the program should be reviewed to assess its readiness for evaluation. Brief discussions of these two steps follow.

**Developing a Logic Model**

Developing a clear conceptualization of prevention and intervention sequences is integral to tracking program progress and grounding the evaluation design in a firm understanding of program goals. Assumptions and hypotheses about the linkages between program activities and program goals are made explicit by developing a logic model. The content of the logic model is developed by discussing goals and the rationale behind program organization and content with service providers and funders, examining planning documents and program reports, and reviewing research findings on similar programs or problems. The literature review may be particularly helpful in identifying theories about the outcomes and any factors other than the program which should be considered in the evaluation.

The logic model begins with a simplified description of the program, the intended outputs, and the intended outcomes. Careful definition of the program components is essential to understanding the program. This includes identification of the types and levels of service elements and may be expanded during development of the design to include measures of quality as well. Output refers to

---

immediate program products resulting from the internal operations of the program, such as the delivery of planned services. Examples of outcome indicators are the numbers of children immunized, home visits by case managers, or youth completing a job training program. Program outputs are, in turn, the vehicle for producing the desired program "outcomes," such as decreases in childhood illnesses, decreases in abuse and neglect cases, or increases in youth employment. It is sometimes useful to categorize outcomes as intermediate versus long term. The developmental stage and expected progression of the youth towards outcomes often dictates careful attention to the time at which the outcome should be expected to occur. For example, improved school attendance in early grades might be an intermediate outcome associated with the longer term outcome of dropout prevention. A classic failure in anticipating this problem occurred in evaluations of the DARE drug prevention program, an educational program for fifth and sixth graders designed to prevent drug use. Evaluation results showed no significant prevention of drug use at the end of the program. However, this result should have been anticipated in view of the fact that drug use does not typically begin among youth in this country until the mid-teen years (14 to 17). An age-appropriate intermediate outcome should have been selected as the primary outcome measure.

The logic model should also include antecedent variables (i.e., pre-existing characteristics of the program organization, community or target population) likely to influence the outputs and outcomes and/or use of program services. The model may also include mediating variables (i.e., conditions or events in the program, target population, or community) that may limit or expand the extent to which program outputs actually produce the desired outcomes. For example, a drug abuse prevention program may be less effective if the program staff are inexperienced, or if the local community offers fewer recreational alternatives to substance abuse and/or more active open drug markets. Alternatively, offering other support services in combination with the program may enhance its impact.
In impact evaluation, the logic model is used to specify the hypotheses, identify the key concepts to be measured, and plan the analysis. The logic model is used to spell out how, and for whom, certain services are expected to create specific changes/benefits. For example, if the program includes parenting classes, the logic model will identify this activity as a key program component and show the types of changes in parenting which will be used to measure program outcomes (e.g., by improving parental assistance with homework or helping parents communicate more effectively with adolescents). Logic models are constructed to show temporal sequence, building left to right, and often diagram relationships with arrows.

In performance monitoring, the logic model is used to focus on which kinds of output and outcome indicators are appropriate for specific target populations, communities or time periods. For example, among indicators of child improvement in school, one might expect attendance to improve in the first semester of a program, but academic test score improvement only after a significant period of program participation -- with the timing possibly varying by the age and developmental stage of the children.

In process evaluation, the logic model is used to identify expectations about how the program should work -- an "ideal type" -- which can then be used to assess the deviations in practice, why these deviations have occurred and how the deviations may affect program outputs. This assists program managers (and evaluators) to identify differences (including positive and negative unintended consequences), consider possible mechanisms for fine-tuning program operations to align the actual program with the planned approach, or re-visit program strategies to consider alternatives.

An example of a logic model is shown in Exhibit A. It was developed by The Urban Institute during the planning of the evaluation of the Children at Risk

---

Program (CAR). CAR is an intensive intervention program designed to prevent involvement in drugs and crime, and to foster healthy development among adolescents ages 13 to 15 who exhibit serious risk indicators and live in severely distressed inner-city neighborhoods.

The intervention consists of eight required program components: 1) Case Managers employed by the program make a service plan for all members of the household of participating youth and provide intensive follow-up on referrals to needed services, handling a caseload of 15; 2) Family Services include parenting skills training for all parents, and referral to other services as needed (intensive family counseling, stress management/coping skills training, identification and treatment of substance abuse, health care, job training and employment programs, housing, and income support services); 3) Education Services include tutoring or homework assistance for all youth, and referral to other services as needed (educational testing, special education classes); 4) After-School and Summer Activities for all CAR youth include recreational programs and life-skill/leadership development activities, combined with training or education; 5) Mentoring is provided by local organizations for youth in need of a caring relationship with an adult. The role of the mentor is to: (a) inform youth about alternative available choices (e.g., activities and goals); (b) familiarize them with strategies available for pursuing those choices; (c) provide training, opportunities for practice, and feedback in the development of skills for implementing particular strategies; and (d) provide relationships through which youth are affirmed, inspired and encouraged to make healthy choices; 6) Incentives such as gifts and special events are used to build morale and attachment to the prosocial goals of the program (e.g., gift certificates, trips, and vouchers for pizza, sports shops, movies and stipends for community service during summer programs); 7) Community Policing/Enhanced Enforcement to create safer environments with less drug activity have been added to all target neighborhoods. Law enforcement activities include out-stationing police in schools and neighborhood locations to maintain order and enhance
relationships with community groups; 8) Criminal/Juvenile Justice Intervention involves collaboration between CAR case managers and juvenile court personnel to provide community service opportunities and enhanced supervision of youth in the justice system.

These program components are designed to achieve reductions in risk factors and enhancement of protective factors at the end of program participation - the intermediate outcomes. These intermediate outcomes, measured at the end of program participation, are hypothesized to be requisite steps towards the desired longer-term outcomes -- prevention of drug use, drug selling, delinquency, school failure and dropout and teen parenthood. The antecedent variables include the levels and types of neighborhood, family, peer group and personal risk factors for participants as well as their demographic characteristic. These are influences that are present before the program intervention. Mediating variables include exposure to other social or educational services, perceptions of opportunities and social norms. These are influences that operate at the same time as the program is hypothesized to have its intermediate effects. Program outputs, not shown in this diagram, will be measured as part of the evaluation. These include indicators of performance such as the number of tutoring sessions provided, number of home visits by case managers, and number of times parents participated in program activities.
Reviewing Readiness for Evaluation

Evaluability assessment is a systematic procedure for examining whether program evaluation is justified, feasible and likely to provide useful information. Questions to be considered in an evaluability assessment include:

- Does the program's logic appear to adequately lead to the program's planned outcomes? If program goals are unrealistic or the intervention strategies not well grounded in theory, then evaluation is not a good investment.

- What kinds of data will be needed, from what number of subjects, and what data are likely to be already available? Evaluations should be designed to maximize the use of available data, as long as these are valid indicators of important concepts and are believed to be reliable. Available data may, for example, include government statistics, individual and summary agency records and statistics and information collected by researchers for other studies.

- What resources and assets are available — money, time, expertise, community and government support? Are there any factors that limit or constrain access to these resources?

- Can the evaluation be achieved in a time frame that is likely to permit the findings to be useful in making program and policy decisions by federal, state and local officials?

- To what extent does evaluation information already exist somewhere on the same or a closely related intervention? The answer to this question can have important implications for action. Any successful previous attempts may yield promising models for replication. Lessons learned from previous unsuccessful attempts may inform the current effort. If sufficient evidence already exists from previous efforts, the value of a new evaluation may be marginal.

- To what extent are the findings from an evaluation likely to be generalizable to other communities, such as by expanded use of the program intervention being evaluated? Are there unique characteristics of the projects to be evaluated that might not apply to most other projects?

---

5 For more information on deciding when and how to make decisions on whether and how to conduct a program evaluation, see Schmidt, Scanlon, and Bell (1979). Evaluability Assessment: Making Public Programs Work Better and Wholey, J. (1994) Assessing the Feasibility and Likely Usefulness of Evaluation.
Selection of the evaluation design follows the systematic consideration of these questions.

Selecting an Evaluation Design

There are six kinds of evaluation designs: 1) experimental designs, 2) quasi-experimental designs, 3) non-experimental designs, 4) performance monitoring, 5) case studies and ethnography, and 6) cost studies. The first three are used to answer questions about the impact of a program. They share the strategy of comparing outcomes with and without program participation. Performance monitoring looks at outputs and outcomes, but does not include comparative data on what would have happened without the program. Case studies and ethnography are often used to address process evaluation questions. Cost studies are usually conducted in combination with an impact evaluation, although performance monitoring can address related questions of program efficiency, effectiveness, and productivity. Each of these designs is described, followed by a diagram illustrating how the most powerful design can be identified given the structure of the intervention and availability of comparison groups.

Experimental Designs

Key elements. Experimental designs are considered the “gold standard” in impact evaluation. Experiments require that individuals or groups, such as classrooms or schools, be randomly assigned to one or more groups prior to the start of services. One or more “treatment” groups may be designated to receive new services designed to achieve clearly specified outcomes. If multiple treatment groups are designated, the outcomes for the treatment groups may be compared to each other to estimate the relative impact of the different services. Usually, the design will include a “control” group which receives no services with similar goals. The treatment group outcomes are compared to control group outcomes to estimate the total program impact. Because chance alone determines who receives the program services, the groups can be assumed to be similar on all characteristics that might affect the outcome measures.
except the program, and alternative hypotheses about the causes of outcomes can be rejected.

**Design Variations.** One design variation is based on a random selection of time periods during which services are provided. For example, new services may be offered on randomly chosen weeks or days. A version of this approach is to use "week on/week off" assignment procedures. Although not truly random, this approach closely approximates random assignment if client characteristics do not vary systematically from week to week and is often easier for program staff to implement than making decisions on program entry on a case-by-case basis. A second design variation is a staggered start approach in which some youth (or classrooms or schools) are randomly selected to receive services with the understanding that the remainder will receive services at a later time (the next semester or month). One disadvantage of the latter design is that the observations of outcomes is limited to the period between the time the first group completes the program and the second group begins. As a result, it is generally used to assess gains made during participation in relatively short-term programs.

**Limitations/Considerations.** Although experiments are the preferred design for an impact evaluation on scientific grounds, random assignment evaluations are not always the ideal choice in real life settings. One major consideration is whether an experiment is feasible. Some interventions are inherently impossible to study through randomized experiments. Youth curfews, for example, cannot be enforced against randomly selected children but not others. Random variation in curfew enforcement levels may not be maintainable for a long enough period to achieve an impact. And "week on/week off" enforcement is likely to breed contempt for both the law and enforcement. In such situations, one of the quasi-experimental designs described below may be the strongest achievable designs.

A second consideration is whether random assignment is ethical and acceptable to the community. Public opinion may resist "social experimentation," "treating similar children differently on the basis of a coin flip." Some individuals
may view random assignment as exploiting vulnerable populations and powerless people. A reasonable procedure for randomization can overcome such sources of resistance in some situations. When resource constraints preclude offering an intervention to all youth in a community, drawing lots is nearly as traditional as "first come first served" as an allocation method. Providing services for some youth at a later time (the next month or semester as described above) may satisfy community concerns about fairness and be consistent with available staff and resources. Sometimes, random assignment can involve relaxing a requirement instead of adding one, which makes randomization less controversial. Great care needs to be taken to ensure that the control group is not denied essential services, that the benefits to participants and the community are carefully explained, and program staff and participants understand and support the research. Many funders require a formal review of the research design by a panel trained in guidelines developed to protect research participants.

Another topic to be considered is whether the potential results justify the investment. Experiments typically require high levels of resources -- money, time, expertise, and support from program staff, government agencies, funders and the community. Thus, evaluation planners have to ask themselves whether the answers to the list of evaluation questions and pending decisions on program continuation, expansion or modification could be based on less costly, if less definitive, evaluation strategies.

**Practical Issues.** Experimental evaluations present substantial difficulties, given the setting of most programs designed for vulnerable children and youth, and requires substantial investment in the research. An experimental evaluation needs to include close collaboration between the evaluation team and the program staff in identifying objectives, setting schedules, dividing responsibilities for record-keeping and data collection, making decisions regarding client contact, and sharing information on progress and problems. Active support of the key program administrators, on-going staff training and communication via meetings, conference
calls, or e-mail are essential.

One problem often encountered in experimental evaluations is failure to adhere to the plan for random assignment. Staff are often intensely committed to their clients and will want to base program entry decisions on their perceptions of who needs, or will benefit from, the program. Whenever possible, the evaluator, not program staff, should be in charge of the allocation to treatment or control group. Statistical adjustments in the analysis may be needed if there are operational failures to maintain the randomization process. And even these may be inadequate to remove the biases thus introduced.

Another potential problem area is sample attrition. Program staff can often collect data and provide contract information for treatment group members because they have continuing contacts with clients, other agencies and the community. Collecting comparable data and contact information on control group members can be difficult. If the treatment and control groups suffer differential attrition, bias can result. The best way to avoid differential attrition is to plan tracing procedures and data collection at the start of the evaluation, gathering information from the control group members on how they can be located, and developing agreements with other community agencies, preferably in writing, for assistance in data collection and tracing. These agreements are helpful in maintaining sample continuity in the face of staff turnover at some agencies.

Other potential problems may require use of special statistical techniques. These include insufficient or unequal follow-up periods for sample members, and the risk of events (e.g., school failure, incarceration, injury, residential mobility) that are more likely to remove some subjects than others from a sample before the

---


end of the planned follow-up period. If the program services and content change over time, it may be difficult to determine what level or type of services produced the outcomes. The best strategy to use is to identify key changes in the program and the timing of changes as part of a process evaluation and use this information to define the "types of program" variations in the program experience of different participants during the impact analysis. Planning for the random assignment usually needs to begin before the program starts so that agreements are in place with all agencies on how services will be allocated. The duration of the evaluation varies depending on the length of the intervention and the timing of outcome observations (program end or some later period).

Examples. The evaluation of the Children at Risk Program, described above, was based on an experimental design in which youth in five cities were identified as meeting the risk criteria for services by school, courts or case managers and asked to participate in an experiment in which they had a 50/50 chance of being selected for the program. Those who agreed were randomly assigned to a treatment or control group prior to the start of services. Private, in-person interviews were conducted with the youth and their primary caregivers at home before the start of services, and two years later at the end of the program period. Youth were interviewed a third time one year after the program period. School records provide data on grades, promotion rates, dropout rates and achievement tests scores; police and court records provide data on detected illegal activities; and program records provide data on the level and type of participation in CAR services. Data from the CAR program was used to measure the level and type of program participation. Program impact was assessed by comparing outcomes for the two groups on interim outcomes such as reduction in family risk and increased attachment to school, and to longer term outcomes such as grade promotion, drug use, and school dropout. The fact that the evaluation includes programs from

---

different regions of the country and participants from different ethnic groups increases the extent to which the results can be generalized to high risk youth in many areas.

Evaluation of this type is costly and in the case of the CAR evaluation exceeded $2,000,000. The evaluation began at the time the program started and will continue for a six year period to allow time to measure outcomes for one full year after participation by the final cohort of youth entering the sample. Multiple visits to each of five sites were required to arrange for random assignment, data retrieval by agencies, and collection of information on program implementation. Data were collected annually from 15 public agencies in five cities. When completed, the evaluation will include approximately 2,000 in-person interviews in the home conducted by professional interviewers. Because the sample is highly mobile and hard-to-locate, extensive tracing efforts are required to attain acceptably high (85%) response rates in repeated surveys. The benefit from this investment will be a rigorous test of a number of practical and theoretical questions about services for high risk youth, their relationship to youth and family characteristics, and the effects of reducing risk in various domains. Preliminary results based on the program’s first year showed significant reductions in police and court contacts, significant increases in promotion to the next grade, improved family organization and reduced family conflict.

A randomized field trial was used to evaluate Project Alert, an eight-week junior high school curriculum for teaching seventh grade students to avoid drug use.\(^9\) Thirty California and Oregon schools were randomly assigned to three groups: 1) students instructed by adult health educators, 2) students instructed by older teenagers, and 3) a no-treatment control group, although four of these schools provided other drug prevention instructional programs. To increase the generalizability of the findings, the schools were drawn from eight urban, cities.

---

suburban, and rural communities and nearly a third of the schools had minority populations of 50% or higher. To enhance the pre-treatment equivalence of groups and enhance the statistical power of the analysis based on a relatively small sample of schools, each experimental condition included at least one school from each community and restricted the allowable school assignments to achieve minimal imbalance among groups on school test scores, language spoken at home, drug use among eighth graders, and the ethnic and income composition of the school catchment area. These procedures produced substantial pretreatment equivalence in factors related to drug use among the experimental groups.

Statistical corrections were used to correct for the clustering of students within schools. Students completed questionnaires about their drug use seven times between grades 7 and 12; those who transferred to other schools or districts completed mail and telephone interviews to minimize sample attrition. Outcome measures included cognitive risk factors associated with drug use: beliefs about consequences of use, norms regarding drug use, resistance self-efficacy, and expected future drug use. The rigorous design permitted strong conclusions about the long term effectiveness of drug prevention education during early adolescence and demonstrated that results are not restricted to middle class communities, but can be used in schools with a high proportion of lower income and minority students. Because the evaluation began before the curriculum was provided and continued to measure outcomes five years after program completion, a considerable period of time was required.

**Quasi-Experimental Designs**

**Key Elements.** Like experiments, quasi-experimental evaluations compare outcomes from program participants to outcomes for comparison groups who do not receive program services. The critical difference is that the decision on who receives the program is not random. Comparison groups are made up of members of the target population as similar as possible to program participants on factors that could affect the selected outcomes to be observed. Multivariate statistical
Design Variations. Non-equivalent comparison groups are made up of non-participants who resemble program participants on need for services, risk factors, family and community setting, personal characteristics, and access to services and resources. Usually, evaluators use existing groups of youth -- those who live in a similar area, or are enrolled in the same school in a different classroom, or attended the same school with the same teacher in the previous year. In some situations, staff (or schools or communities) are willing or trained to try the new “treatment” while others are not, but the same rules for service eligibility are used by all. The primary alternative design is to identify individuals for the comparison group by matching them to individuals in the treatment group on a selected set of characteristics. This process for selecting a comparison group is methodologically less defensible.\(^\text{10}\)

Quasi-experimental designs also vary in the number and timing of the collection of data on program outcome measures. The selection of the number and timing of measurements is based on an assessment of the potential threats posed by competing hypotheses. In many situations, the strongest designs are those that collect pre-program measures of outcomes and risk factors and use these in the analysis to focus on within-individual changes that occur across the program period. These variables are also used to identify groups of participants who benefit most from the services. One design variation involves additional measurement points to measure trend more precisely. Another -- which is useful when pre-program data collection such as administering a test on knowledge or attitudes might “teach” youth about the questions to be asked -- involves limiting data collection to the end of the program period for some groups.

Considerations/limitations. The use of non-equivalent control group designs requires careful attention to procedures that rule out competing hypotheses.

regarding the causality of observed differences on the outcomes of interest. In
evaluations of programs for vulnerable children and youth, three threats to validity
stand out.\(^{11}\)

The first is the threat of "maturation" which refers to the possibility that
age-related processes will contribute to outcomes independently of the program
intervention. Among youth, certain outcomes, positive and negative, are strongly
tied to age -- outcomes like drug use, delinquency and early parenthood. It is
therefore necessary to be sure that the comparison group is made up of youth at
the same developmental stage.

A second threat is that of "history." This refers to the risk that unrelated
events may affect outcomes. For example, the rapid spread of crack use among
women of childbearing age in the United States in the late 1980s greatly increased
rates of drug-exposed infants. Thus, a comparison group for an evaluation of a
prenatal health care program would need to be drawn from the same years and
communities to "control" for the spread of crack. Otherwise, the upward trend in
negative outcomes due to crack could obscure the prevention benefits of the
program. Similarly, designs need to consider controls for geographic variation in
events external to the program. For example, localized gang crackdowns in some
neighborhoods and not others could influence assessments of the impact of a
school-based delinquency or drug prevention program. If the crackdown occurred
in the "treatment" neighborhood, the program effects might be over-estimated; if it
occurred in the comparison neighborhood, program effects might be under-
estimated.

A third threat to validity is the process of "selection" -- the factors which
determine who receives services. Some of these factors may be easily identified
and used as control variables in statistical models, such as living in a specific
school district or meeting program eligibility criteria. However, it is unlikely that all

Chicago: Rand McNally.
factors will be correctly identified and adequately measured. For example, program participants may receive services because they are more motivated, skillful, or socially well-connected than nonparticipants. Such differences are not easy to measure during a program evaluation.

Practical Problems. Building defenses or "controls" for threats to validity into evaluation designs through the selection of comparison groups and the timing of outcome observations is a creative challenge. Controls for maturation, history and selection may involve, respectively, selecting a sample that includes multiple age cohorts, collecting data in similar or near-by localities that lack the program or applying a statistical model that controls for foreseeable biases in selecting program participants. Even when the comparison group was carefully selected, the researcher cannot be sure that all relevant group differences have been identified and measured accurately. Statistical methods can provide adjustments for unmeasured variables, unequal probability of selection, and differences between the groups at both the contextual and individual level using hierarchical linear modeling. Despite increases in the precision with which program effects can be estimated using these techniques, they do not fully compensate for the non-random design and findings still need to be interpreted carefully and untested alternative hypotheses considered.

As in experimental evaluation, plans for quasi-experimental evaluations need to pay close attention to the problem of collecting comparable information on control group members and developing procedures for tracing them. However, the

---


need for close collaboration with program staff is reduced, since the staff are generally not involved in selecting participants and may have had no contact with comparison group members. Quasi-experimental evaluations, like experimental evaluations, should start before the services are provided with a plan which specifies which subjects or groups will receive services and which will serve as the comparison group. However, it may be possible to begin recruiting comparison groups a program has started.

Examples. A quasi-experimental design was used to evaluate comprehensive health care for high risk adolescents provided by seven clinics funded by the government to provide a wide range of services focusing on the special needs of adolescents. Outcomes for youth served by the seven funded clinics were compared to youth served by three clinics that did not offer specialized services for adolescents. The sample included clinics from different cities and regions of the country. The funded clinics linked medical school training programs in adolescent medicine to services and were divided into two groups: 1) those located in hospital or neighborhood clinics, and 2) those located in school. Samples of consecutive adolescent clinic admissions were selected. The adolescent patients were interviewed at the time of service, usually at the clinic, and 12-months later, usually at home. This data was linked to data abstracted from their medical records for 2,788 patients, 72% of those originally selected for the sample. This design was selected because it was not possible to randomly assign hospitals or youth to adolescent health care linked to medical schools. To control for differences among the patients, the analysis controls for number of presenting problems, age, sex, and race. The evaluation examined the impact of comprehensive care by examining reductions in the number and type of behavioral and lifestyle problems as well as the frequency of sexually transmitted disease and pregnancy during the year between interviews. This strong quasi-experimental

---

design was able to control for characteristics of the service delivery setting and client characteristics at the start of service. However, the study could not fully control for differences in medical risk because to a large extent problem recognition and diagnosis is a function of the skill of the staff, which was likely to be higher at the specialized clinics. This may account in part for the finding of no significant differences between the outcomes reported by youth served by these kinds of clinics.

The evaluation of the Teen Age Parenting Program (TAPP) for adolescents divided teen mothers into three groups designed to be similar in age and other characteristics.17 Each group was one-third black, hispanic and white. One group attended an alternative school with child development and parenting classes and a nursery school which featured a parenting-child development curriculum. Another group attended an alternative school without a nursery school; the remaining group received no special services for teenage parents. Services began during pregnancy. Assessments of educational progress, fertility, knowledge, and child development two to four years later were based on interviews and school records. Mothers in the alternative school with the nursery program had completed more schooling and were more likely to still be enrolled in school than the other mothers. Mothers in both alternative school programs had more knowledge about parenting and reproduction and more positive attitudes about parenting than those without special services. However, there were no significant differences in the groups on child development outcome measures. However, because the design did not have pre-program measures of individual differences and assignment was not random, these significant differences in outcomes may be due to selection bias, with more promising cases assigned to receive special services.

Non-Experimental Evaluations

Key elements. Non-experimental evaluations examine changes in levels of risk or outcomes among program participants, or groups including program participants, but do not include comparison groups of other individuals or groups not exposed to the program.

Design Variations. The four primary types of non-experimental designs include: 1) before and after comparisons of groups of program participants, 2) time series designs based on repeated measures of outcomes before and after the program for groups which include program participants; 3) panel studies based on repeated measurement of outcomes on the same group of participants, and 4) post-program cross-sectional comparisons of participants.

The first two designs are based on analysis of aggregate data. In before and after comparisons, outcomes for groups of participants (program cohorts) are measured before and after an intervention and an assessment of impact is inferred from the differences. This simple design is often used to assess whether knowledge, attitudes or behavior of the group changed after exposure to a classroom curriculum or job training program. Time series designs are an extension of the before and after design that uses multiple measures of the outcome variables before an intervention begins and continues to take multiple measures after an intervention was put into place. The hypothesis that a change in the trend (in direction or level) in the outcomes occurred at, or shortly after, the time of the intervention is tested statistically. Time series measures may be based on larger groups or units that include program participants. For example, crime rates for neighborhoods in which most or all youth receive a delinquency prevention program might be used to assess reductions in illegal activity. Evaluation of a series of dropout prevention activities offered across the school year could examine the percentage of entering classes reaching graduation over a period of years. Time series designs are most useful when it is difficult to identify who received program services or when the evaluation budget does not support collection of
detailed data from program participants. Although newer time series statistical techniques permit the inclusion of covariates in the statistical analysis, it is difficult to rule out the impact of non-program events using this approach.

The next two designs examine data at the individual level. Cross-sectional comparisons are based on surveys of groups of participants conducted after program completion. This design can be used to estimate correlations between outcomes and differences in the duration, type, and intensity of services received. Panel designs use repeated measures of the outcome variables for each individual. In this design, the outcomes are measured for the same group of program participants, often starting at the time they enter the program and continuing at intervals over time. For example, the evaluation of Health Planning and Promotion: Life Planning Education used pre-post data from participants to show measurable gains in understanding the best combinations of contraceptive methods and the consequences of early childbearing. This design allows the characteristics of individual participants to be used in the analysis to identify different patterns of change associated with individual characteristics of participants and control for other events to which they were exposed.

Considerations/Limitations. Several limitations to non-experimental designs should be noted. First, the panel designs and cross-sectional provide only a segment of “dose-response curve” estimates of the marginal differences in impact related to differences in the services received. These designs cannot estimate the full impact of the program compared to no service at all unless estimates can be based on other information on the risks of the target population. Second, the designs that track participants over time (before and after, the panel design, and time series) cannot control for the effects of developmental changes that would have occurred without services, nor for the effects of other events outside the

---


program’s influence. Third, the extent to which the results can be assumed to apply to other groups or other settings is limited because this design provides no information for assessing the extent to which participants were selected into the program on the basis of factors which determine the outcomes.

Practical Issues. Non-experimental designs are usually relatively easy and inexpensive to conduct. Individual data for cross-sectional or panel analysis is often collected routinely by the program at the end (and sometimes beginning) of program participation. When relying on program records, the evaluator needs to review the available data against the logic model to be sure that adequate information on key variables is already included, or to begin collecting additional data items if needed.

When individual program records are not available, aggregate statistics may be obtained from the program or from other community agencies with information on the outcomes among groups of participants. For example, crime rates, average promotion rates, and rates of births to teen mothers can be collected from existing records. The primary problem encountered in using such statistics is that they may not be available for the specific population or geographic area targeted by the program. Often these routinely collected statistics are based on the general population or geographic areas served by the agency (e.g., the police precinct or the clinic catchment area). The rates of negative outcomes for the entire set of cases included may well differ from rates for the targeted group of vulnerable children and youth; this risk is greater for larger statistical areas than for smaller areas.

A more expensive form of data collection for non-experimental evaluations is a survey of participants some time after the end of the program. These surveys can provide much needed information on longer term outcomes such as rates of employment or earnings or high school graduation. As in any survey research, the quality of the results is determined by response rate, rather than overall sample size, and by careful attention to the validity and reliability of the questionnaire.
items. One practical advantage of non-experimental evaluations is that they can be conducted after a program has been completed and can compile data for many years through retrospective data collection procedures. Thus, outcomes across many years can be studied in the time required to assemble and analyze the data (as little as a year).

**Examples.** A non-experimental design was used to evaluate the San Diego Youth and Community Services Youth Development Project which provided four levels of service to high risk youth in junior high school. The service levels were: early identification and school-based prevention (primary prevention), two levels of intervention with youth who have committed a crime or are homeless (secondary prevention), and services for multi-problem youth who have been found by the court to have committed a crime or have been homeless for an extended period of time (tertiary prevention). Core services, assigned on an individual basis as needed by case managers, included education, vocational and pre-employment training and placement, leadership and independent living skills, counseling and recreation. Pre-post comparisons of participants were used to assess the extent to which adolescents in the program made gains towards project objectives of staying free of criminal activity and living self-sufficiently. Data came from intake assessments conducted by case managers, the case manager's narrative assessment of progress while in the program, ongoing assessments administered by program staff during program participation at six month intervals and a follow-up assessment at 24-months by a professional evaluator. In addition, juvenile probation records were used to document arrests and probation violations during the program and 24-months later. The evaluation examined outcomes against pre-defined program goals. For example, 60% of the participants were expected to show gains in self-concept, and 80% will show improvements at 24 months on their educational, vocational and social functioning and self-sufficiency. However, the absence of a comparison or control group makes it impossible to assess whether the outcomes resulted from program inputs, characteristics of the participants or features of the
context. The costs of the evaluation were not separated from the costs of program operation, but can be expected to include the costs of professional evaluators to design the instruments for pre-post data collection, conducting the follow-up survey and the analysis and report preparation, as well as the program staff time used for assessments.

The Youth Training Scheme (YTS) in Great Britain provides, through local agents, two years of vocational and on-the-job training for out-of-school and unemployed youth ages 16 and 17. The local agents are businesses or community organizations which receive government funds to design a training program, recruit and supervise youth, and provide at least 13 weeks of on-the-job training per year. Non-experimental evaluation of YTS was based on a follow-up survey of 63,000 former participants. In addition to monitoring client satisfaction and job related outcomes, the survey was used in non-experimental comparisons of differences in outcomes related to differences among participants: job market outcomes were compared for graduates versus program dropouts and across youth who entered the program with different levels of motivation and past school achievement. Results indicate that program graduates had better labor market outcomes than those who did not complete the program. Similarly, earning qualifications in the program (an interim outcome measure) was positively correlated with later labor market success (the longer term outcome). Non-experimental comparisons were also used to identify differences in outcomes related characteristics of the participants or the training experience. The field of employment and type of local agent providing the training were significant predictors of labor market outcomes. Similarly, labor market outcomes were better for youth who began the program with higher levels of motivation and past school achievement. However, these findings highlight the shortcomings of non-experimental design: these same youth might have been more likely to become

---

employed in the absence of any program participation.

**Performance Monitoring**

**Key Elements.** Performance monitoring is used to provide information on: 1) key aspects of how a system or program is operating; 2) whether, and to what extent, pre-specified program objectives are being attained (e.g., numbers of youth served compared to target goals, reductions in school dropouts compared to target goals); and 3) identification of failures to produce program outputs, for use in managing or redesigning program operations. Performance indicators can also be developed to:

1) monitor service quality by collecting data on the satisfaction of those served, and

2) report on program efficiency, effectiveness; and productivity by assessing the relationship between the resources used (program inputs) and the output and outcome indicators.

If conducted frequently enough and in a timely way, the process can provide managers with regular feedback that will allow them to identify problems, take timely action, and subsequently assess whether their actions have led to the improvements sought. Performance measures can also stimulate communication about program goals, progress, obstacles and results among program staff and managers, the public and other stakeholders. They focus attention on the specific outcomes that are desired and better ways to achieve them, and can promote credibility by highlighting the accomplishments and value of the program.

Performance monitoring involves the identification and collection of specific data on program outputs, outcomes and accomplishments. The data are numeric, although they may measure subjective factors such as client satisfaction and may consist of frequency counts, statistical averages, ratios or percentages. Output measures reflect *internal* activities: the amount of work done within the program or organization. Outcome measures (immediate and longer term) are measures of amount or level of progress towards program goals. Often the same
measurements (e.g., number/percent of youth who stopped or reduced substance abuse) may be used for performance monitoring and impact evaluation. However, unlike impact evaluation, performance monitoring does not attempt to distinguish whether these were caused by program efforts or by other external events.

**Design Variations.** When programs are operating in a number of communities, the sites are likely to vary in mission, structure, the nature and extent of project implementation, primary clients/targets and timeliness. They may offer somewhat different sets of services, or have identified somewhat different goals. In such situations, it is advisable to construct a "core" set of performance measures to used by all, and to supplement these with "local" performance indicators which reflect these differences. For example, some youth programs will collect detailed data on youth school performance, including grades, attendance, and disciplinary actions, while others will simply have data on promotion to the next grade or whether the youth is enrolled or dropped out. A multi-school performance monitoring system might design one set of indicators to be used by all which would require data on promotion and enrollment, and more detailed, or specialized, indicators on attendance or disciplinary actions for one or a subset of schools to use in their performance monitoring.

**Considerations/Limitations.** In selecting performance indicators, evaluators and service providers need to consider:

- **The relevance of candidate measures to the mission/objective of the local program or national initiative.** Key considerations are: 1) do process indicators reflect program strategies/activities identified in mission statements? 2) do outcome indicators cover objectives identified in mission statements? and 3) do indicators capture the priorities at the community level?

- **The comprehensiveness of the candidate set of measures.** Does the set of performance measures cover inputs, outputs, and service quality as well as outcomes and includes relevant items of customer feedback?

- **The uniqueness of each candidate measure.** To what extent does each measure duplicate or overlap with other candidate indicators?
• **The program’s control over the factor being measured.** Does the program have influence/control over the outputs or outcomes measured by the indicator? If the program has only limited influence over the outputs or outcomes being measured, the indicator may not fairly reflect program performance.

• **The validity of the measure.** Do the proposed indicators have content validity — do they reflect the range of outcomes the program hopes to affect? Is the data free from obvious reporting bias?

• **The reliability and accuracy of the measure.** Can indicators be operationally defined in a straightforward manner so that supporting data be collected consistently over time, across data gatherers, and across communities? Do existing data sources meet these criteria?

• **The feasibility of collecting the data.** How much effort and money is required to generate each measure? Should a costly measure be retained nonetheless because it is perceived as critically important?

**Practical Issues.** The set of performance indicators should be simple. It is best to use a limited number of key indicators of priority outcomes. Too many indicators burden the data collection and analysis and make it less likely that managers will understand and use reported information. At the same time, the set of indicators should be constructed to reflect the informational needs of stakeholders at all levels — community members, agency directors and national funders.

Regular measurement, ideally quarterly, is important so that the system provides the information in time to make shifts in program operations and to capture changes over time. Thus, the timing of this type of evaluation is best if it begins with the start of program operations and continues throughout. However, pressures for timely reporting should not be allowed to sacrifice data quality. For the performance monitoring to take place in a reliable and timely way, the evaluation should include adequate support and plans for training and technical assistance for data collection. Routine quality control procedures should be established to check on data entry accuracy and missing information. At the point
of analysis, procedures for verifying trends should be in place, particularly if the results are unexpected.

The costs of performance monitoring are modest relative to impact evaluations, but vary widely depending on the data source. Most performance indicator data come from records maintained by service providers. The added expense involves regularly collecting and analyzing these records, as well as the expense of preparing and disseminating reports to those concerned. This is typically a part-time work assignment for a supervisor within the agency. The expense will be greater if client satisfaction surveys are used to measure outcomes. An outside survey organization may be required for a large scale survey of past clients; alternatively, a self-administered exit questionnaire can be given to clients at the end of services. In either case, the assistance of professional researchers will be needed in preparing data sets, analyses, and reports.

Examples. A number of programs serving vulnerable youth and children provide data to document their performance. Rio’s Community-Based Adolescent Health Clinic: Adolescent Health Unit reports that over 25,000 health visits by youth were recorded at the central clinic. This indicator is a measure of program output -- services delivered. The clinic also reports that 74% of 182 participants in a survey reported a change in their behavior as a result of services received. This indicator measures a program goal directly and thus is an example of an outcome measure. The percentage of youth in the target population (youth in the favela in which the clinic is located) reached by clinic services increased from 3% to 14%. This measure of the extent to which the program accomplished its targeted goals has the advantage of supplying a standard, albeit somewhat arbitrary standard, against which to judge program performance in the absence of a control or comparison group.

The Asociacion Salud con Prevencion (ASCP) in Colombia, a non-

---

Examples of good performance measures vary by the type of program. Indicators for early childhood development programs such as Head Start include the percentage of participants scoring at or above grade level on standardized reading achievement tests and the percentage of participants immunized against childhood diseases, drawn from program records. Indicators for drug prevention programs include the percentage of participants who report that no drug use during the year after the program and the percentage who believe drug use is harmful. These indicators require self-report data from participant surveys or end-of-program questionnaires, administered under conditions of privacy. Dropout prevention programs frequently use the percentage of participants who advance to the next grade or increases in the overall promotion rate of a targeted school from school records. Indicators of the performance of teen pregnancy prevention programs include the percentage of births or pregnancies among participants based on program or medical records, the decline in teen birth rate in a targeted area from governmental organization (NGO) which provides primary prevention services which promote adolescent reproductive health, monitors outputs with data on the number of professionals trained, the number of youth given educational services, the number of workshops held, the number of condoms distributed, and the number of medical and counseling sessions provided. The results demonstrate that the program is providing promised services, but does not give an indication of the impact in terms of either immediate outcomes such as use of birth control or longer term outcomes which include reduced risk of out-of-wedlock births or early childbearing. Overall annual operating costs for ASCP were reported, but not by service category or number of adolescents served. This reflects the difficulties broad-based, primary prevention programs have in defining who receives program benefits and how to allocate program activity costs to multiple service components. In contrast, programs such as Benosta - Nacion de los Muchachos, an NGO that provides residential programs for Colombian youth no longer living with their families, reports costs per youth.
vital statistics, or the percentage of participants who use birth control or abstain from sexual activity, based on self-report data collected under conditions of privacy.

**Case Study/Ethnography**

**Key Elements.** Case studies involve the detailed analysis of selected program sites or clients to determine how the program is operating, what barriers to program implementation have been encountered, what strategies are the most successful and what resources and skills are necessary. The answers to these questions are useful in providing guidance to policy makers and program planners interested in identifying key program elements and in generating hypotheses about program impact that can be tested in the impact analysis. The results are used in the program evaluation to:

1) judge program implementation by determining whether the model is being implemented as specified, and, if not, how operations differ from those initially planned; and 2) revise the logic model continually as new understandings of program operations and key elements emerge. Case studies are sometimes used to test competing hypotheses about differences in the impact of services. This approach is often the only feasible one when competing models have emerged in different locations and the desire is to assess which is most successful in attaining goals shared by all. This requires purposely selecting sites to represent variations in elements or types of programs, careful up-front analysis of potential causal models, and the collection of qualitative data to elaborate the causal links at each site.

Clients or sites chosen for case studies should represent wide variation in settings, program models and clients. The identification of sample members within sites, interview topics, and key data elements begins with the logic model as a guide. In a case study, qualitative data, collected using semi-structured interviews and observations of program operations, are supplemented and verified by collecting quantitative data on program operations and performance from records.
Ethnography is a research design which relies almost exclusively on qualitative data collected to study:

- Organizational and programmatic processes occurring at a program site;
- The community context in which the program is taking place;
- The relationship between program activities and other activities in the community;
- Causal processes as the participants view them; and
- Modes of decision making.

Ethnography does not begin with the logic model, but starts by trying to understand the program from the perspective of staff, participants and others in the community. Ethnographers observe program operations as unobtrusively as possible, sometimes in the role of participant observer, and keep detailed field notes which are transcribed and coded to identify emerging themes and trends. The critical research goal is to provide data on the subjective experience of those in the program situation and to use this information to understand how, and if, the program goals are being achieved.

*Design Variations.* Case studies may use several different approaches for collecting qualitative data for program evaluation. The most frequently used are semi-structured interviews, focus groups, and researcher observations while on-site. Semi-structured interviews allow for the discovery of unanticipated factors associated with program interpretation and outcomes. Protocols for semi-structured interviews contain specific questions about particular issues or program practices. The "semi" aspect of these discussion guides refers to the fact that a respondent may give as long, detailed, and complex a response as he or she desires to the question -- whatever conveys the full reality of the program's experience with the issue at hand. If some issues have typical categories associated with them, the protocols will usually contain probes to assure that the researcher learns about each category of interest.

Focus groups also are structured forms of inquiry involving two researchers, a facilitator and a recorder. Usually five or six general questions are selected to
explore issues of how the program is perceived and subjective understandings of program strategies or goals. These are used to guide an open-ended discussion lasting about an hour and a half. The goals of the discussion may vary from achieving group consensus on some points to emphasizing points of divergence among participants. Discussions are tape-recorded, but the primary record of the focus group is the detailed notes taken by the researcher who is acting as recorder for the group. Less detailed notes may also be taken publicly, on a flip-chart for all to see, to try to achieve consensus and give group members the chance to add anything they think is important. Soon after a focus group, two researchers (recorder and facilitator) summarize in writing the main points that emerged in response to each of the general questions. When all the groups are completed, the researchers develop a combined summary, noting group differences, and hypotheses about those differences.

In case studies, observations at program sites provide an important method of validating information from interviews. In this case, the observations will often be guided by structured or semi-structured protocols designed to ensure that key items reported in interviews are verified and that consistent procedures for rating program performance are used across time and across sites.

Ethnography is based on observations that are much less structured and uses procedures that are deliberately flexible. As a result, ethnography is helpful in gathering information on unintended consequences and unanticipated outcomes. These unexpected observations may lead to an entirely new concept of program delivery. In a recent project examining service integration programs for at risk youth, observations helped clarify that service integration needed to go beyond formal links and on-paper agreements, and provided insights into how informal processes bonded services together in their efforts to make a difference for high risk youth in the community. At the same time, observations conducted as part

---

of anthropological or ethnographic studies are perhaps the hardest type of qualitative information to analyze, since they tend to generate volumes of information, much of which may not be directly related to evaluation goals and may not be comparable across sites.

**Practical Issues.** Collecting qualitative data requires skilled researchers who are experienced with the techniques being used. To analyze these data, careful notes must be taken to ensure that responses are correctly recorded and to aid in interpreting the responses. In methods based on interviews, interviewers must be trained to understand the intent of each question, the possible variety of answers that respondents might give, and ways to probe to assure that full information about the issues under investigation is obtained.

Analysis of qualitative data requires an in-depth understanding of programs, respondents and responses, and especially the context in which they are evaluated. Ultimately, the analyst makes judgments regarding the relative importance or significance of various responses and, therefore, must assess responses in as unbiased a manner as possible to determine whether they support or refute hypotheses about the way the program works and the effects it has.

One way to handle the qualitative data gathered during evaluations is to treat one’s interview and observational notes as text, and to conduct a textual analysis using specialized computer software that can search for the presence of specific themes or content. Qualitative software such as **ASKSAM** or **Ethnograph** are examples of such software. These programs facilitate the location and retrieval of information from massive textual files, and may be useful if the researcher has no other, easier and quicker, way to retrieve and organize the information. However, they may be expensive to use because huge amounts of text must be entered into a computer. Further, either the exact words one wants to search for must appear in the text, or the text be marked for the presence of any theme or topic that the researcher wants to retrieve. Often researchers can achieve equal or better results with carefully constructed interview or data collection guides or
carefully structured focus groups, careful attention to recording responses or coding data encountered in the field, and sensitive interpretation of the assembled results. Qualitative studies must be conducted during program operations. Retrospectively asking participants or staff about their understandings and attitudes is subject to a number of biases, such as poor recall and the distortion of intervening experiences.

**Examples.** An evaluation of mentoring in the juvenile justice system conducted by Public/Private Ventures was based on case studies of two pilot programs. The program was designed to match 100 mentors to at-risk youth. Mentors were trained to meet with youth one-on-one before and after the youth’s release from juvenile detention facilities with the goal of establishing a meaningful attachment to an adult role model. Data were collected from mentor logs, program records, court records, structured interviews with mentors and youth before and after program participation, staff interviews, focus groups with mentors, youth and service agency staff, and in-depth interviews with mentor-youth pairs. The analysis was primarily a qualitative examination which examined the characteristics of successful matches, issues in program implementation, the style and content of mentoring interactions and program staffing. Although it does not offer evidence of the outcomes of mentoring for youth development, the evaluation provides extremely useful information on the process of implementing a mentoring program and guidance for program development and replication.

Proyecto Atlantida, a qualitative research project supported by the private foundation -- Fundacion FES -- used individual and group interviews and narrative research to study the youth culture in small, medium and large cities in Colombia. Researchers and field coordinators from eleven universities working with youth interviewers used ethnographic and case study methods. The study investigated the subjective views of youth of the agencies and institutions, such as the family,

---

which have traditionally supported youth socialization into productive adult roles. The results identified key areas in which Colombian culture needs additional support to help with appropriate youth socialization -- thus offering guidance on the targeting and content of programs for this population. For example, the study found that parents have lost their influence with youth and peer influences have become increasingly important. A strong and identifiable youth culture has emerged that emphasizes consumerism and rapid change. Socialization is increasingly taking place in public places such as school, although schools are seen as irrelevant for many, if not most youth. Like most case study and ethnographic research, the study does not address issues of program impact and only tangentially address other evaluation questions about how services are perceived by clients and what characteristics might increase program appeal and relevance. However, in times of rapid cultural change such as the speed with which urbanization is occurring in many countries, an understanding of the cultural changes is needed in planning programs which will be accepted and used. Costs of this research were not reported, the costs of such a labor-intensive data collection and analysis effort are likely to be quite high relative to more structured evaluation designs.

**Cost Studies**

**Key Elements.** Cost studies are used to assess the investments in programs for high risk youth by collecting information on: 1) direct program expenditures; 2) the costs of staff and resources provided by other agencies or diverted from other uses; 3) costs for purchased services; and 4) the value of donated time and materials. Costs for the first two items usually include expenditures for staff salaries; fringe benefits; special training costs (if any); travel; facilities; and supplies and equipment that have to be purchased. The value of donated resources, which can be substantial, generally has to be estimated and requires careful documentation of the donation. Cost analyses indicate that donations are a major
cost item in many youth programs. For example, the Cities in Schools\textsuperscript{24} evaluation indicated that donations are between 74% and 90% of the total direct program costs, and that the wide variation among cities in the types of donations received made the inclusion of these costs essential to an understanding of the resources required to sustain program operation.

Total program costs and an average cost per client calculated by dividing the total by either the total number of clients served, or the total number of clients who meet some standardized definition of success. This type of cost calculation is then linked to results of an experimental or quasi-experimental impact evaluation to estimate costs per successful client or used with performance indicators to assess the cost or cost-efficiency of achieving program goals.

A second approach to cost estimation calculates the cost per unit of service. For example, the cost per hour of classroom instruction or the cost per hour of counseling. This type of cost calculation is then used in impact evaluations (including non-experimental evaluations) to look at the costs of different outcomes. This type of cost analysis is difficult in multi-faceted, comprehensive programs in which the level and type of service are highly variable and may involve a number of service providers. It is also difficult in programs in which defining exposure to services is difficult. Where possible, it is preferable to distinguish between fixed costs (e.g., rent or the director's salary) and variable costs (e.g., the costs of special events or the hourly costs of the recreation director). The variable costs can then be used to estimate the marginal cost of adding additional clients to the number receiving a specific unit of service.

**Design Variations.** Cost studies can be undertaken to describe the program costs and link these to the level of outcomes achieved. In this application, the costs are compared to the level and type of outcomes documented in performance monitoring outcomes. Decisions on whether the outcomes justify the costs are

based on opinions about the value of the outcomes (not monetized) and the likelihood that the outcomes were attributable to the program.

Cost-effectiveness analysis is used to compare the costs of different approaches to providing some standard level of service or desired level of outcome. This approach is most useful when multiple programs are using different models to provide a service. The requirements are that the programs serve the characteristics of target populations served, the program goals, and the output or outcome measures be identical. For example, cost-effectiveness studies could compare the relative effectiveness of residential and non-residential treatment for drug abusing youth, provided that the services were provided to youth who were similar in age and drug use problems and the same measures of treatment success applied to both programs.

Cost-benefit studies provide estimates of the dollar benefits returned for each dollar spent on the program. This is the key question from a policy perspective, but is not easily answered. This type of evaluation has rigorous requirements for: 1) an estimate of program costs, either per client or per unit of service; 2) estimates of the value of the benefits, and 3) comparative data on program impact – an estimate of outcomes with and without the program. The first item should be obtainable from program financial records, supplemented as needed by estimates of the cost of donated or reallocated resources. The second can be obtained from an experimental or quasi-experimental evaluation of program impact or another strategy for estimating what might have happened without the program. The primary barrier to conducting cost-benefit analysis of programs for vulnerable children and youth stems from the third item: difficulty of placing dollar values on benefits. Many benefits appear to be of intrinsic value, for example, reductions in family dysfunction and conflict, but do not lend themselves to monetizing. For example, the evaluation of Rio’s Community-Based Adolescent Health Clinic: Adolescent Health Unit defends the higher costs of health services for adolescents provided by a specialized unit, on the grounds that the unit helps
guarantee higher quality services and a focus on the special needs of this population -- benefits for which monetary values are not easily estimated.

Monetization of benefits to individuals requires assumptions about three matters; all are frequently controversial. The first is the dollar value of the benefit which may depend on personal values, for example, what residents are willing to pay for a crime-free neighborhood. The second is the time discount rate: a dollar of benefit today is worth more than a dollar benefit realized next year. The third is who benefits. For example, a high school equivalency degree for a violent youthful offender may result in the same gains in lifetime earnings for the offender as a violence victim would realize from physical therapy for the injury. Social values become important when the beneficiaries differ in standing and perceived merit. To circumvent such difficult questions, the analyst may conduct a sensitivity analysis to reach conclusions based on explicit assumptions of value. For example, the neighborhood crime prevention program may be deemed cost-effective if "residents are willing to pay at least $100 per month for 10% lower rates of burglary" or "if the discount rate is less than 6%" or "if the offender's earnings are worth 50% of the victim's earnings."

Beyond benefits to individuals, the total value of benefits includes the social costs averted. These are the savings to the public which result from avoiding negative outcomes. The values, therefore, are based on studies which document the social costs of negative outcomes such as the costs of crime or drug abuse. These estimates are difficult to derive and are often based on tenuous assumptions. To compensate for problems in the reliability of the value estimates, cost-benefit calculations normally use a range of benefit values in an attempt to place an upper and a lower bound on the probable returns to investments in the program. A more significant problem is that monetary values based on public costs for the negative outcomes among the general population may be poor

estimates of the value of benefits among the program’s target population. For example, national estimates of the costs of drug abuse may not apply to reductions in amphetamine abuse among low-income adolescents in a single city. This problem needs to be acknowledged and value estimates revised to the extent possible to reflect savings for the program’s participants. Other public benefits reflecting gains, not costs averted, are widely acknowledged, but rarely find their way into cost-benefit studies. Examples include improvements in the quality of life or the environment. The lack of public consensus on the priority to be attached to improvements in areas such as these is a barrier to estimating values.

Considerations and Limitations. Documentation of gains to prevention programs is exceptionally difficult and requires estimating negative outcomes that did not occur. As described above, the most robust estimates of program impacts of this kind are based on experimental evaluations, which are difficult and expensive to conduct, or on quasi-experimental evaluations, which are slightly less difficult and expensive. When the program has total population coverage, it is possible to interpret differences between the observed trend and predicted trend in an outcome indicator over time to program impacts and estimate the monetary value of the benefits. This strategy was used to estimate the value of drug prevention efforts in the United States. National survey estimates of drug use in 1979 were used to estimate expected drug prevalence during the 1980s and early 1990s; the differences between these estimates and drug use prevalence rates based on national surveys during these years were attributed to Federal investments in drug prevention programs.28

Practical Issues. The issue faced in most program evaluations is that developing a conceptual framework that reflects all the issues in benefit valuation, and then devoting the resources necessary for estimating the range of benefits, can require as much or more research time and expertise than determining whether

the program had any impact. However, evaluating program impact is usually the
top priority since there inevitably is a limit on research dollars and benefit values do
not matter in the absence of program impact. A number of studies of the value of
preventing negative outcomes among children and youth have been initiated
recently. These can be expected to give program evaluators much greater access
to estimates of the value of reductions in youth problems for use in cost benefit
studies in the future.

Cost studies are based on retrospective data on program costs and must
wait for accounting to be completed for the period of time selected for study
(which generally excludes the start-up period). When costs must be estimated from
observations or diaries which record program activities and staffing, the data
collection must take place during the period of operations selected for the cost
study. However, analysis of cost benefit and cost effectiveness must wait until
data are available from the impact evaluation estimating the type and magnitude of
program outcomes and thus often take place in the last year of an impact
evaluation or even in the following year.

Examples. Evaluation of the Jobs Program, a preventive intervention for
unemployed persons, examined the costs and benefits of the program from three
perspectives, that of: the participants, the program funder (the Federal
government) and the government of the state in which the workers lived. 27 The
analysis used outcome data from a randomized field experiment to estimate the
monetary value of differences in income for participants and non-participants,
projecting earnings to age 60 in current dollars based on data on earnings, time
worked, number of jobs and other labor market outcomes collected one month,
three months, and 32 months after the program. The results indicated a
significant, enduring and relatively large difference in income between participants
and non-participants estimated at $239 per month at the 32 month follow-up

27 Vinokur, A. D., vanRyn, M., Gramlich, E. M. and Price, R. H. (1991) "Long-Term Follow-up and
Benefit-Cost Analysis of the Jobs program: A Preventive Intervention for the Unemployed." Journal of Applied
point. Program costs of training program staff and implementing and providing services, including overhead, were estimated to be $286 per participant. The gains were substantial from all three perspectives. The individual participant was expected to benefit by $5,392 at the end of 32 months, by at least $10,377 at the end of five years, and by at least $38,944 by age 60. Benefits to the federal government in the form of increased tax revenues for the same periods were estimated to be at least $720, $1,649, and $6,979, while those to the state were estimated to be at least $308, $12,619, and $48,151 at a 5% discount rate. These gains were averaged over all enrollees; returns to those who completed the program (did not dropout) were substantially higher. State costs averted in the form of unemployment benefits not paid to participants was omitted from the calculation because the difference in the number of days eligible for unemployment compensation, which required a waiting period, was not significant between participants and non-participants.

An evaluation of 13 delinquency prevention programs in Los Angeles County estimated cost effectiveness as a function of the delinquency risk of the population of youth served, costs, and success rate. This study compared cost to benefit ratios of alternative programs designed with a common goal and outcome measure -- preventing subsequent arrest. Because the risks of delinquency varied among the youth served by different programs, estimates of the risk of delinquency was derived from existing research and used to classify the youth served by the program into four risk categories. Program costs were estimated by taking the total budgets from all sources divided by the number of clients. Costs of public expenditures for delinquency (costs to the community and justice system) were estimated from the proportion of the justice system budget (from the County budget) devoted to juvenile cases, divided by the number of juvenile cases at various stages of processing (from annual reports of the Los

Angeles Probation Department, the California Youth Authority and the U.S. Department of Justice).

The public costs averted were calculated in the same way by dividing the budget by the number of arrests of youth following program participation and calculating the savings as the difference between the two. In addition, the benefits of reductions in expected future arrests were estimated based on the probability of subsequent arrests reported in studies of criminal careers times the estimated public savings per arrest averted. Savings to victims were based on estimates of the costs of damage and loss for each type of juvenile offense from earlier research, adjusted for inflation. These costs per offense were applied to the expected lifetime arrests in the absence of the program and benefits estimated to be the difference between these costs and the absence of costs associated with no further arrests or victimization (estimating that for each arrest, there are four to five offenses that do not result in arrest). Thus estimated program benefits were the sum of the public costs averted and the savings to victims.

The results were used to estimate the cost differential (costs divided by the value of benefits) to programs with different rates of success (measures as arrests prevented), controlling for the risk of offending of the juvenile population served. The findings were used to estimate the success rate required to show a positive rate of return given the delinquency risk of the population served for programs with different cost differentials. This then can be used in monitoring the performance of a wide variety of delinquency prevention programs.

Choosing among the Impact Designs

Choice of an impact evaluation begins by identifying the design that offers the strongest capacity for isolating the independent causal effects of the program and is feasible given the structure of program. The "decision tree" shown in Exhibit B illustrates a process for identifying which alternatives are feasible.

If the program will be provided to a limited number of youth who can be identified in advance and randomly selected for participation, then an experimental
design should be considered. If the program will be provided to a limited number of youth, but those who receive services is determined by organizational or geographic considerations (or other non-random selection rules), then quasi-experimental design variations should be considered.

The most difficult design challenges occur when the program is intended to serve all members of the target population. If the new program is implemented fully and rapidly, no youth will be available for a comparison group. Often, however, new full-coverage programs -- for example, new health services -- are intended for an entire population, but are not implemented in every community in the country, and certainly not at the same time. If some communities or groups are not included in the implementation, it may be possible to select as comparison sites some communities that did not implement the program and use a quasi-experimental design. However, this may not solve the problem of comparability sufficiently to allow such a design if the communities where it was implemented have unique characteristics that prompt them to offer to use the program. Only if it is possible to select randomly the location or participants can an experimental evaluation of the program which rules out competing hypotheses be conducted.

Non-experimental designs may be the best available choice when a full-coverage program is not implemented uniformly. If a program is implemented at different levels across sites but uniformly within sites, one would choose a cross-sectional design. But if a target population is exposed to different levels of the program within a community, one would choose a panel study design, follow a sample of individuals, and record both outcomes and the amount of the program or intervention each individual received and when it occurred. If defining who is served by the program is difficult or the program is uniformly applied in all communities, then a time-series design would be appropriate. Before-and-after designs without control groups are often used, but are subject to a number of threats to validity, including maturation and secular changes.
Identifying Potential Evaluation Problems

Evaluations of youth programs also face a number of challenging problems in applying research methods to the kinds of programs that serve vulnerable children and youth. A summary of these, based on experience in reviewing and evaluating programs for these populations, is provided below to guide development of realistic evaluation plans.29

Defining program participation. Programs for youth may be open-ended, lacking formal intake procedures and policies for determining when the program was “completed.” Counselors may be contacted for several chats, followed some weeks later by an appointment, followed by intermittent participation in some, possibly not all, of the services offered. Youth may stop attending and then resume. Limiting participation in the evaluation to those who attend regularly may create a “selection bias” by dropping from consideration the youth who are most difficult to engage. Often identifying who “participated” and for how long requires multiple categories to reflect the variations in type, duration and intensity among the youth served. Participants should be followed from the point of first contact and all major program activity documented. Evaluators also need to decide whether others who potentially benefit from the program, such as parents, boyfriends/girlfriends, or siblings, are program participants. If so, their participation in program activities should also be tracked; if not, plans need to be made on how to count the gains made by these indirect program beneficiaries in evaluating program impact.

Evaluating the relationship between participation and outcomes. Many programs emphasize individualized services in which youth with the highest levels of risk are offered the greatest number or most intensive level of services. Obviously, assignment to treatment in this case is not random, and the higher

problem youth may never achieve the same level of positive outcomes as youth who began with fewer problems. For example, studies of the School-Based Health Centers in the U.S. show that frequent clinic users were at greater risk for alcohol and substance use, sexual activity, and poor family and peer relationships. Thus, comparing their outcomes to those for nonusers or those who used the clinic less frequently would be inappropriate. Similarly, comparisons between different youth programs must consider the type and level of risk exhibited by participants and this information used to identify who is helped by the program. For this reason, data on the risks and needs of participants should be collected at intake for use in analysis and a pre-post design used when possible.

Defining the unit of analysis. Deciding on the appropriate unit of analysis can be difficult, particularly in evaluating comprehensive programs. Programs may target entire neighborhoods, classrooms, or families for change -- sometimes planning activities directly for different groups, and at other times planning carryover effects. Measurement at multiple levels is frequently appropriate and can be used in combination with analytical techniques such as hierarchical linear modeling. For example, crime reduction can be assessed by comparing neighborhood rates of calls for police services, household victimization rates or youth delinquency surveys. Economic gains can similarly be measured by changes in the unemployment rate, average household or family income and individual earnings. The selection should be closely linked to program goals and activities.

Evaluations of services integration programs, including most that use a case management approach, will face additional challenges in: 1) tracking the services received by participants; 2) developing common agreements among agencies on program goals and required components; 3) measuring effects on the service delivery system; and 4) differentiating services integration from service comprehensiveness.

---

Tracking the services received by participants. Services integration usually involves referring participants to other agencies for needed assistance. A critical, and often difficult, problem is determining which services were actually received. Clients may or may not contact agencies to which they are referred, they may or may not be accepted for services, and they may or may not participate in services. Documenting the chain of participation is essential to determining the extent to which services integration is being achieved, but is time consuming and often resisted by programs who see making the referral as the extent of their responsibility. Because staff turnover in service agencies is frequently high, preparing written agreements on data access and sharing is strongly recommended. In the absence of adequate documentation, information on service utilization can be collected in follow-up interviews with clients.

Developing common agreements among agencies on program goals and required components. The agencies collaborating in a services integration effort may differ in their vision of the program's goals, the key strategies to be used, and in how youth needs will be evaluated and problems addressed. Evaluations tend to highlight these differences and may encounter difficulties in gaining consensus. This is particularly true when multiple agencies recruit clients and/or case management services are not centralized. Time should be allocated for face-to-face meetings to get agreement on whom evaluators will count in selecting measures of program outcomes, and how service provision is expected to achieve program goals.

Documenting service delivery by multiple agencies. In services integration programs, many agencies coordinate and combine their resources to meet the needs of clients. One of the most difficult problems can be assembling information on who received what types and amount of service. Agencies have different methods of identifying clients -- some use family identification numbers, others identify individual children served. Some group service records by family or child; others maintain records by contact, which introduces multiple records for single
clients which then have to be unduplicated. Agencies such as schools or juvenile
courts sometimes face legal or professional barriers to sharing client-based
information with other agencies or evaluators. A systematic system for collecting
the data needed to compile a complete picture of program participation needs to be
developed early in the planning process and, as noted above, be supported by
written agreements and on-going technical assistance and staff training in record-
keeping procedures.

Measuring effects of the service delivery system. A primary goal of services
integration is to change agency operations and increase effectiveness. These
outcomes need to be measured at the agency, not individual, level. Evaluations of
services integration need to document changes in agency procedures, increased
participation in collaborative planning and service delivery, and decreases in
barriers to interagency cooperation and client service associated with policies, and
procedures. Referral patterns should show more diversity in planning. At the
individual level, clients should report fewer unmet service needs, shorter waiting
periods for service, and increased satisfaction with the response to their needs.
Other evidence of integration includes increased staff knowledge and familiarity
with the resources of other agencies and community groups.

Differentiating services integration from service comprehensiveness.
Services integration is intended to give youth and children in need not only faster,
more appropriate services, but also services they would not otherwise obtain. The
referral process educates clients on the options and assistance potentially
available. Improved interagency planning and coordination reduces the barriers to
obtaining additional services. This makes the task of differentiating services
integration from service comprehensiveness very difficult. The evaluation and
program staff need to develop clear expectations on the extent to which the ease
of obtaining services and the appropriateness of the service package can be
distinguished from the extent to which the program is providing comprehensive
services to meet the full range of client needs.
Conclusions and Recommendations

Strong pressure to demonstrate the impact of programs for vulnerable children and youth dictates making evaluation activities a required and intrinsic part of program activities from the start. At the least, the evaluation activities should include performance monitoring. The collection and analysis of data on program progress and process builds the capacity for self-evaluation and should contribute to good program management and efforts to obtain support for program continuation. When the funding is expected to serve as "seed" money to start a program and demonstrate its effectiveness in the hope that its activities will continue under local sponsorship, this kind of evaluation is the essential minimum. Performance monitoring can be extended to non-experimental evaluation with additional analysis of program records and/or client surveys. These evaluation activities may be conducted either by program staff with research training or by an independent evaluator. In either case, training and technical assistance to programs to support evaluation efforts will be needed to maintain data quality and assist in appropriate analysis and use of the findings.

There are several strong arguments for evaluation designs that go further in documenting program impact. Experimental or quasi-experimental designs are needed as a basis for convincing arguments that program funds are well invested, and that the program is making a contribution to the well-being of the children and youth served. These evaluations need to be conducted by experienced researchers and supported by adequate budgets. A good strategy may be implementing small-scale programs to test alternative models of service delivery in settings that will allow a stronger design than is possible in a large scale, national program. Often program evaluation should proceed in stages, with the first year of program operations devoted to process studies and performance monitoring serving as a basis for supporting more extensive evaluation of a program once operations are running smoothly.

The final recommendation is that planning be extensive around the issue of
obtaining support for the evaluation at every level – community, program staff, agency leadership and funder. Each of these has a stake in the results and should have both a voice in planning, and clearly defined benefits to flow from the results. Only in this way will the results be valid and applied.
EXHIBIT A. Logic Model Used in Evaluation of the Children at Risk Program

Target Population
- 11-13 years old
- Live in high crime
- Area
- Attend target school

Resources
- $ 
- Labor
  - paid
  - Volunteer
- Equipment

Type and Level of Program Inputs
- Case management
- Educational assistance
- After school/summer programs
- Mentoring
- Incentives
- Enhanced drug enforcement
- Community policing
- Juvenile justice intervention

Intermediate Outputs
- Reductions in Risk Factors
- Enhancement of Protective Factors

Outcomes
Prevention of:
- Drug use
- Drug dealing
- Delinquency
- School failure/drop-out
- Teen parenthood

Antecedent Variables
Risk level and type:
- Neighborhood
- Family
- Peer Group
- Personal
Demographic Characteristics:
- Age
- Ethnicity
- Family Composition
- Caregiver characteristics

Mediating Variables
- Level and type of utilization of services outside the program
- Perception of opportunities
- Educational aspirations
- Social norms and attitudes
EXHIBIT B. Process for Selecting Impact Evaluation Designs

1. New Study
2. Random Assignment OK?
   - Yes
   - No
3. Near-random alternative feasible?
   - Yes
   - No
4. Find Comparison Group
5. Interrupted Time Series Available?
   - Yes
   - No
6. Time Series Available?
7. Uniform Implementation Status?
   - Yes
   - No
8. Non-uniform
9. Micro Data Available?
   - Yes
   - No
10. Panel Design
11. Aggregate Cross-section Design
12. Non-equivalent Comparison Group Design w/ Statistical Controls
13. Known Selection Process?
   - Yes
   - No
14. Adjust Model for Selection Bias
15. Select a Design

Prepared by Dr. Jeffrey A. Roth, The Urban Institute, 1995
Other Reports in the LASHC Series

No. 1: Children of the Poor in Latin America and the Caribbean by J. van der Gaag and D. Winkler, July 1996 (available in Spanish translation)

No. 2: Targeting At-Risk Youth: Rationales, Approaches to Service Delivery and Monitoring and Evaluation Issues by M. Schneidman, July 1996 (available in Spanish translation)

No. 3: Evaluating Programs for Vulnerable Children and Youth by A. Harrell, July 1996 (available in Spanish translation)

No. 4: Chile’s Learning Network by M. Potashnik, July 1996

No. 5: Review and Analysis of International Experience with Programs Targeted on At-Risk Youth by G. Barker and M. Fontes, July 1996 (available in Spanish translation)

No. 6: Measuring Public Hospital Costs: Empirical Evidence From The Dominican Republic by M. A. Lewis, May 1995