DOCUMENT RESUME

ED 274 438 PS 016 058

AUTHOR	Barnard, Kathryn; And Others
TITLE	Program Evaluation: Issues, Strategies and Models. A Discussion for Administrators and Practitioners Serving Disabled and At-Risk Infants, Toddlers, and Their Families.
INSTITUTION	National Center for Clinical Infant Programs, Washington, DC.
PUB DATE	86
GRANT	MCJ-113271-01-1
NOTE	22p.
AVAILABLE FROM	National Center for Clinical Infant Programs, 733 Fifteenth Street, N.W., Suite 912, Washington, DC 20005 (Free).
PUB TYPE	Viewpoints (120) -- Reports - Descriptive (141)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Action Research; Clinics; *Disabilities; *Family Programs; Formative Evaluation; *Hospitals; Infants; Models; Parents; Program Descriptions; *Program Evaluation; Summative Evaluation; *Toddlers
IDENTIFIERS	*Early Intervention; *Infant Care

ABSTRACT

Offering a discussion for administrators and practitioners serving disabled and at-risk infants, toddlers, and their families, these three addresses explore issues, strategies, and models of program evaluation. "Major Issues in Program Evaluation," by Kathryn Barnard, discusses the purpose of, process of, and ways of doing program evaluation, as well as the development of a model relevant to a wide range of early intervention programs. "Evaluation Strategies," by Victoria Seitz, gives three evaluation strategies: (1) process evaluation, which describes a project's client population and activities; (2) outcome evaluation, which tries to ascertain whether people are different as a result of receiving service; and (3) action research, which explores the mechanisms by which a successful intervention program helps those for whom it does work. Seitz also explores a variety of innovative research designs with which to evaluate intervention outcomes. "The Infant-Parent Program at San Francisco General Hospital," by Jeree Pawl, relates the process by which staff of a service program for young children and families worked together to design and implement an evaluation process which helped to improve clinical practice and supervision while simultaneously demonstrating the efficacy of the program to funding sources. The report also summarizes discussion elaborating themes raised by the three major speakers and relates program evaluation issues to current concerns of state administrators and community practitioners. (RH)

Program Evaluation:
Issues, Strategies and Models

NATIONAL CENTER FOR CLINICAL INFANT PROGRAMS

Program Evaluation:
Issues, Strategies and Models

A discussion for administrators and practitioners serving disabled and at-risk infants, toddlers and their families

NATIONAL CENTER FOR CLINICAL INFANT PROGRAMS

An exploration of issues based on a meeting sponsored by Project Zero to Three A Special Project of Regional and National Significance supported by the Division of Maternal and Child Health, Grant NO. MCJ 113271-01-1, published by the National Center for Clinical Infant Programs, 1986.

Foreword

"We've had the demonstration projects. Now how can we take what we've learned and get services to all disabled and at-risk children in our state who need them—starting from birth?"

"Our state's system of perinatal care is saving newborns who never would have survived a few years ago. But how can we make sure these babies get key follow-up care? How can the school system keep track of them and plan to meet their needs when they start school years from now?"

"How can we convince state legislators that early intervention is effective? What do we say when they ask what early intervention means? What do we mean by "effective?"

Practitioners, parents and policymakers across the country who are concerned with the needs of disabled and at-risk infants, toddlers and their families are struggling with a set of questions like these, questions related to the broader issues of how we can establish permanent programs for this vulnerable population. There is growing commitment to serve handicapped children from birth to three. But this commitment must be reflected in support for services that remain in place year after year, services that families can count on. Such an ongoing system must reflect accurately the need for services, ensure that agencies live up to delegated responsibilites, and include a mechanism for evaluating both the integrity of service programs and the quality of services offered in the light of new research in the field.

Because administrators at the state level are frequently in the lead in examining these issues, the Division of Maternal and Child Health began in 1983 to fund the National Center for Clinical Infant Program's Project Zero to Three, an initiative which brought together ten states (Maine, New Jersey, Maryland, North Carolina, Ohio, Iowa, Texas, Utah, Washington and Hawaii) well along in the process of building comprehensive service systems for disabled and at-risk children in the first three years of life. As representatives from these states voiced their concerns, plans were made to hold three meetings during the summer of 1984 to discuss:

1) Legislative mandates for serving disabled and at risk children from birth to three—What are the advantages, barriers, pitfalls and alternatives?
2) Systems for identifying and tracking high-risk (high priority) infants—What kinds of systems are most useful? How can improvements be made?
3) Program evaluation—What are the purposes of assessing the effectiveness of programs? How are these goals best accomplished?

"Program Evaluation: Issues, Strategies and Models" grew out of a meeting of administrators, researchers and clinicians which took place in Columbus, Ohio in July, 1984, under the sponsorship of the National Center for Clinical Infant Programs' Project Zero to Three with support from the Division of Maternal and Child Health. Two additional publications, "There Ought To Be A Law?: Ensuring Statewide Services for Disabled and At-Risk Infants and Toddlers" and "Keeping Track: Tracking Systems for High-Risk Infants and Young Children" were products of similar meetings.

Thanks for ongoing, invaluable encouragement and support for this publication, as well as for every aspect of Project Zero to Three, go to Vince Hutchins, Merle McPherson and Camille Cook of the Division of Maternal and Child Health, Bureau of Health Care Delivery and Assistance. Special appreciation is due to Kathy Peppe and Betty Macintosh who shouldered primary responsibility for arranging the Ohio meeting; to Kathryn Barnard, Jeree Pawl and Victoria Seitz, who were featured speakers and discussion leaders; and to participants from Hawaii, Ohio, Texas and Washington who generously shared their ideas, experiences and concerns.

Summary

Although evaluating programs of services for disabled and at-risk young children and their families is a complex undertaking, service providers can design and carry out evaluations which:

- help *practitioners* improve their services and reach currently unserved populations in need.
- help *policymakers* develop realistic expectations for early intervention programs and understand the need to support evaluation efforts at the program level as well as broader, more comprehensive external research projects.

This booklet contains three addresses relevant to issues, strategies and models of program evaluation.

- *Kathryn Barnard*, Ph.D., R.N., Professor at the School of Nursing and affiliate of the Child Development and Mental Retardation Center of the University of Washington, discusses: 1) the purpose of program evaluation; 2) the process of program evaluation; 3) ways of doing program evaluation; and 4) developing a model relevant to a wide range of early intervention programs.
- *Victoria Seitz*, Ph.D., of Yale University, describes three evaluation strategies: 1) process evaluation, which describes a project's client population and activities; 2) outcome evaluation, which asks whether people are different as a result of receiving service; and 3) action research, which explores the mechanisms by which a successful intervention program helps those for whom it does work. Seitz also explores a variety of innovative research designs by which to evaluate intervention outcomes.
- *Jeree Pawl*, Ph.D., Director of the Infant-Parent Program at San Francisco General Hospital, University of California at San Francisco, relates the process by which staff of one service program for young children and families worked together to design and implement an evaluation process which helped to improve clinical practice and supervision while simultaneously demonstrating the efficacy of the program to funding sources.

This report also summarizes discussion among the local service providers, state agency administrators and researchers who participated in the July, 1984 meeting that focused on program evaluation. This discussion elaborated themes raised by the three major speakers and also related program evaluation issues to current concerns of state administrators and community practitioners serving disabled and at-risk infants, toddlers and their families.

Major Issues in Program Evaluation

by KATHRYN BARNARD

I would like to address four major issues involved in program evaluation:

1) the purpose of program evaluation
2) the process of program evaluation
3) ways of doing program evaluation and
4) the development of a model relevant to a wide range of programs for infants, toddlers and their families.

The purpose of program evaluation

Giving guidance to the program is the most important reason to do program evaluation. Service providers need some idea of whether the kinds of services they are giving do, in fact, meet the needs of the target population.

A model of program evaluation which offers guidance is one developed in nursing by Dr. Joanne Horsley of the Oregon Health Sciences University, Portland, Oregon. In her "research utilization model," she suggests that in any service delivery program it is staff which must be involved in the process of promoting change.

- The first step in the process is to identify potential change agents within a staff. These individuals may have a bit more education than their colleagues, may be a little bit more innovative, and often have a "sparkle in their eye."
- Involve these people in a process of reviewing the literature and talking to other programs about some problem of real concern to your own program. If, for example, children's absenteeism from your program is a problem, see if anyone has studied the factors involved.
- Develop a protocol to deal with the problem.
- Test the protocol or approach for six months or a year.
- Look at the results.
- Using those results, move ahead to generalize a successful approach to the problem you're concerned with or to other problems.

In short, one of the ways a service delivery program can use evaluation well is by focusing on a specific aspect of the program, designing something to test and try (a field trial) and getting people involved in evaluating the new approach.

I believe strongly that before one tests a model extensively there should be at least some data to suggest that the model has been effective. To test a completely new model of early intervention at several sites across the country can be a risky undertaking—especially if you find if doesn't work.

> *Giving guidance to the program is the most important reason to do program evaluation. Service providers need some idea of whether the kinds of services they are giving do, in fact, meet the needs of the target population.*

In the field of early intervention—despite the emphasis on individual program plans—we probably do too much generic programming. I am convinced that different children in different types of families need different kinds of programs. Handicapped infants in emotionally unavailable families need different services from those required by handicapped infants in available, resourceful families. Even though the modal problem among a group of children in an early intervention program may be developmental delay or mental retardation, such programs must also meet the needs of children with sensory disorders, physical handicaps, and multiple disabilities. To be able then, to test a model of home-based, center-based, parent-oriented or child-oriented intervention *meaningfully*, one must first have a really clear sense of the specific characteristics of the children and families to be served and their program needs.

In the process of seeking support to become established, early intervention programs may have suggested to parents, pediatricians and others that early intervention can achieve more ambitious goals than it is realistic to expect. What kinds of expectations does your community have about your intervention program? Do people think that there will be a lower rate of children requiring special services in the schools? Do they think that families will be spared divorce and conflict because of early intervention programs? What are the expectations?

We could provide a real service by clarifying expectations about what early intervention programs can deliver. At some period in the history of a program, the funding source may simply ask how many children are receiving service. But other questions will follow. It is important for programs to educate parents, funding and referral sources and political bodies about appropriate evaluation questions, so that in fact programs are asked questions which they can answer in an effective way.

The process of program evaluation

Everybody seems to want programs evaluated—particularly the people who fund them and the people who run them—but my experience has been that evaluations represent extremely threatening endeavors to the people who are actually offering services. As a consequence, service providers are likely not to want the evaluation or even to believe its results.

But since intervention staff will have to get involved in some aspects of the evaluation—reporting how they see children's progess or documenting certain behaviors, for example—an evaluation will be much more effective if it is planned and supported by a majority of program staff. This point is emphasized in United Cerebral Palsy's monograph, *Programming for Atypical Infants and Their Families*, which points out that a theoretically very effective program evaluation plan will fall flat without staff cooperation or people to implement it.

One issue to consider is the extent to which an evaluation will be part of the service program. Will certain clinical indicators be collected as part of the intake process, as part of the ongoing periodic evaluation of children? Will those usual clinical indicators—perhaps with some additions—be part of the evaluation plan? Or will the evaluation be entirely separate from the program? I think the best course is to make the program evaluation the least intrusive to the program, the least burdensome to program deliverers. Evaluation is most likely to get done if this consideration is primary.

For programs new to evaluation, proceeding in stages may be useful: start with some simple indicators, provide feedback to staff, and then continue. One health department with which I am familiar is trying to institute the use of "encounter records" which can become part of a uniform, computer-compatible system. But although the forms to be used are fairly well-conceived and fairly simple, staff are quickly losing interest in doing them, feeling that they are nothing but "number generators." Everyone is so busy getting the system into operation that no one has taken the time to tell the staff about the data emerging from the process: even a count of how many encounters a particular program has had in a month would help to keep the system going. When, however, the people collecting that data hear nothing for six months or a year, their cooperation begins to slip.

In contrast, our hospital perinatal program now has all its statistics computerized and each month can give the staff a printout of how many C-sections were performed, how many twins were born, what medications were used, etc. Obstetricians, perinatologists and nurses are all finding this data extremely instructive in thinking about rates of occurance of certain conditions and in planning in-service training.

I can't emphasize strongly enough the importance of evaluation for improving service delivery. I would hope that any evaluation system has such improvement as a primary goal.

Craig Ramey suggests that measuring the satisfaction of the intervenors is one important part of evaluation. If intervenors do *not* feel that what they are doing is effective, they will leave the field or become "dead wood": thus service programs need to collect information about how satisfied intervenors themselves are with the progress they are making.

An evaluation will be much more effective if it is planned and supported by a majority of program staff. It is important to know how satisfied intervenors themselves are with the progress they are making.

What about the satisfaction of the consumers of our services? One of my colleagues has often said that consumer satisfaction should be one of the most important issues in the delivery of health and educational services. Is it enough that people are satisfied with what they're getting? In Hawaii not long ago, satisfied parents contacted seventy-four out of seventy-six legislators when their early intervention program's budget was at stake: that was convincing data.

If a program does individual program planning, seeing whether the objectives set for individuals are met can be a strong component of an evaluation. I would say that a program developing individual program plans for children and families should aim for an 85% achievement rate. Getting a rate much higher than that may indicate that families and children are not being stretched enough toward individual achievement; going much lower—to 50%—probably means that expectations for the child and family are too advanced. Reporting *collectively* how successfully the program is meeting the goals of *individual* program plans provides a powerful indicator of the program's effectiveness and should be a basic step in the design and modification of the program's intervention effort.

Developing a model relevant to a wide range of programs for infants, toddlers and their families

I have been asked on a number of occasions whether it is realistic for service programs to do evaluation themselves. My answer, basically, is "yes"! We need program evaluation to meet program needs but we also need some projects which can tackle the problems of experimental design, sampling and data collection in order to test theories and models and draw generalizable conclusions.

> *"We should no longer be preoccupied with the question of whether early intervention has a positive effect on IQ."*

We also need to think about a frame of reference for a program evaluation plan. The United Cerebral Palsy monograph to which I have referred lists a number of long-term goals for the disabled child and his family which came out of a meeting of handicapped adults and their parents. These included: 1) independence; 2) a sense of self-worth, including adjustment to the disability, use of talents and abilities and constructive ways of dealing with feelings of helplessness and frustration; 3) integration of the disabled individual into the community; 4) positive integration into the family; 5) parental ability to deal constructively with their feelings about the child; and 6) a parental sense of competence. The monograph also suggests ways of using these objectives to design behavioral observations, interviews, and questionnaires. It might be quite useful for early intervention programs to look at these goals, decide to what extent they might be adopted or modified, and evaluate whether they can be used as a frame of reference for zero-to-three programs.

Ways of doing evaluation

Programming for Atypical Infants and Their Families discusses several evaluation models:

- the psychological model, representing a basic experimental design in which one looks for individual or group differences;
- the educational model, in which individual objectives are set for each client and achievement is measured;
- the objective standard model, which looks at items relevant to the entire client population although not necessarily problematic for all of them;
- the case study model, which includes information about the client, the intervention, and the resultant changes in the client in a narrative account;
- the consumer evaluation model, which asks clients to evaluate the program and its components, and
- the outside evaluation model, in which a person independent of the program staff evaluates the program either in terms of the program's achievement of its stated goals and objectives or in terms of criteria chosen by the evaluator.

I think that one of the biggest mistakes service delivery programs can make in evaluation is to attempt the psychological model, with all the attendant obligations of experimental control. While Victoria Seitz will be helping us look in more detail at alternatives to the psychological model of evaluation, let me suggest that most service programs do not need data which are generalizable for proving to the world that early intervention, in general, is effective. Service programs *do* need data about how effective *they* are with their unique combination of staff, service delivery goals, clients, and community.

We should no longer be preoccupied with the question of whether early intervention has a positive influence on IQ. We need to recognize that a child's IQ is only one of the factors in development. We should be asking what types of services, in what combinations, and following what models are most effective and most cost effective. We should be asking how we can prevent decline in development. We need to ask how we can ensure that all the services that a child and family need are available and provided.

Availability of services

Finally we must think about whether the services we feel are necessary for infants, toddlers and their families are in fact available in our programs and communities. The 1980 Division of Maternal and Child Health publication, *Guidelines for Early Intervention Programs*, talks about protecting the physical, psychological and emotional health of the child and

We need program evaluations to meet program needs but we need some projects which can tackle the problems of experimental design, sampling and data collection in order to test theories and models and draw generalizable conclusions.

family. According to these guidelines, intervention programs for children from birth to three should include at least an assessment of health needs and if possible services to ensure appropriate health care.

In one program, for example, simple physical growth measures revealed that more than 80% of the infants and toddlers enrolled were undernourished. While this program might not have been able to meet the nutritional needs of the children it served, documenting them was certainly important, as was recognizing how strongly physical health problems affect developmental progress, whatever the other intervention approaches used.

Longer range objectives

As we design early intervention programs for disabled and at-risk infants, toddlers and their families and ways to evaluate their effectiveness, we would do well to consider some of the longer-range objectives of our efforts. These might include:

- The absence among our children of treatable, but untreated defects;
- Broader, yet more accurate measures of the status of children from birth to three years of age, including assessment of the caregiving environment;
- The integration of the child, with or without disabilities, into the family so that the child's needs are met, but not at the expense of family satisfaction;
- The feeling of parents that they can handle the situations they face; and
- Decreased use of special services in school and in later life on the part of individuals with disabilities and their families.

Conclusion

Whether one is trying to improve a program or to document its effectiveness and efficiency, program evaluation is important. Our task is to make evaluation a meaningful, integral part of our service delivery programs.

Evaluation Strategies

by VICTORIA SEITZ

Full-time researchers and full-time service providers have a lot to learn from each other and both can gain a tremendous amount from a cooperative relationship. If you leave a full time researcher alone in her ivory tower, not knowing the subject population, not knowing the service program, she'll come up with some pretty silly stuff—and I speak from experience. But a full time researcher and a good program service provider or administrator working together, using the strengths of both, can end up with the best possible outcome evaluation information. I hope this kind of collaboration will be the wave of the future.

After some basic discussion of different kinds of evaluation, I would like to describe alternative research designs for outcome evaluation and then briefly touch on the issue of costs.

The kinds of evaluation, as I see them, are process evaluation; outcome evaluation; and something that has been called action research.

Process evaluation

Process evaluation describes the population the project is serving and what exactly the project is doing. I would like to argue that absolutely anybody who runs a program should do something in the way of process evaluation. At the very least a program should be able to say, for example, "We saw 250 people: we saw most of them for 5 visits; we saw ten of them for 20 visits." This gives *some* kind of picture of what really happened—how many people were seen, for how long, and how intensively.

Process evaluation can also be useful if it describes the people whom the project misses. An outside researcher may have the time to go and look. For example, New Haven has a special alternative school for pregnant teenagers that has been in existence for twenty years. In planning an evaluation of the program and in thinking of ways to avoid selection bias, we decided to try to find every teenager who delivered a first-born child in New Haven during a calendar year and to find out something about her—age, ethnicity, etc. If we knew enough about everybody who *could* have gone to the alternative school, perhaps we could say something meaningful about the ones who actually did.

I must say at this point that it was very hard indeed to find out how many people could have been referred to that school and hadn't been; that is why I don't think that somebody who isn't a full time researcher has time to do this.

> *A full time researcher and a good program service provider or administrator working together, using the strengths of both, can end up with the best possible outcome evaluation information.*

We went to the city hospitals. City hospitals can tell you exactly how old a mother is when her baby is delivered. What they don't have readily available, and what it takes you weeks of sleuthing to find out is whether the mother is a city resident (and this is important, since this school is in the city system.) Is this a first-born child? Has this mother already graduated from high school? In order to find this information, you take your list of hundreds of names and call assorted records offices, which are invariably across town and sometimes administered by a surly person who won't let you have more than ten records at a time. It takes hours to get ten records, and you're finished looking in five seconds—they're not New Haven residents.

The process takes hours of labor. But what we ended up with was some extremely important information, namely that the use of this program depends very much on ethnicity. It turns out that this school successfully reaches 85% of all the black girls in the city of New Haven that it should reach. And since about half the girls it doesn't reach are in fact staying in their regular school, one finds that only about 7%

of the relevant black teens are failing to receive a school program during their pregnancy. However, we found that only 10 to 25% of the pregnant white or Hispanic girls are making it into this school. Armed with this process information, the staff is now making much more active efforts to work with these girls, collaborating with parent schools to develop a system to get many more of these young women at risk involved with the intervention program.

In brief, having basic information about whom you serve and how many of the people whom you would like to serve but whom you are not reaching is useful. If your community has most information computerized, you might be able to get the data you need in a flash. If it isn't easy to find the data, try to find somebody who will do the scutwork for you, because it is worth knowing.

Process evaluation, then, starts with the issue of who is being served and goes on to document exactly what is being done. If you can't describe exactly what your program is doing, how is somebody else going to do the same thing? The very least you can do is prepare a manual which someone could use to replicate the program. Such a description should include not only how many times one see people but how long each session lasts. If staff are on call for crises, make home visits, or follow up actively to make sure referrals succeed, this heavy personal involvement should be documented. It's priceless, and I also suspect that it is very effective. Programs with this level of staff commitment, my instincts tell me, are going to have a real impact on people's lives: if you are doing it, document it.

Outcome evaluation

What providers want to know is: are people different because they are receiving my services? This is not only difficult but practically impossible to find out; but when you do it, the results often show that a program is effective in ways that you never dreamed. If you can find an adequate control group, you can sometimes demonstrate that an amazing amount has been happening—but you can't show it without the control group. Alternative research designs which address this issue will be discussed later.

Action research

Action research asks why a program works or fails to work; it is a creative combination of process and outcome research. Since it is almost always the case that a program works better for certain kinds of people than it does for other kinds, this kind of evaluation allows a program to target interventions better and also to understand the mechanisms which make an intervention succeed.

We are finding, for example, that the greater portion of her pregnancy a girl spends at New Haven's special school for pregnant teenagers, the more likely she is to have a healthy full term baby. It's quite a dramatic effect: 38% of girls who spend no time at all at the school have unhealthy, preterm or low birthweight babies. This percentage drops to about 12% if they have two prenatal quarters, and if they spend their entire pregnancy in the school program, they deliver healthy babies.

Action research asks *why*. Why would a school program have a health effect? Program staff may tell you. First of all the New Haven Hospital's prenatal clinics have established a very nice cooperative relationship with the school. They send their nurses over on a regular basis to talk with the teenagers. What then happens is that the girls like the nurses. And they learn about what prenatal care is, and what's going to happen to them when they come over to the clinic, and how staff really want to see them over there. The nurse gets to know them by name. What tends to happen is that these girls go to the prenatal clinic, know people there and feel much more comfortable. Furthermore, if they *don't* keep getting their prenatal care, there's a lot of peer pressure in the school applied until they do get it.

It is quite clear, then, that the medical mechanism which is giving these young women the prenatal care they need in order to deliver healthy babies is stimulated by what is obviously an effective program on the part of the staff.

Since it is almost always the case that a program works better for certain kinds of people than it does for other kinds, this kind of evaluation allows a program to target intervention better and also to understand the mechanism which makes an intervention succeed.

Action research can also help us answer questions about the population a program is not reaching or for whom if is not working well. In the situation I have been describing, an action researcher would go out and interview some white and Hispanic pregnant teenagers to find out *why* they are choosing not to enter the program. Since one has gotten their medical records, one would also look to see whether there is suggestive evidence that they might be a more troubled group than the black girls.

Ultimately, action research should show us not only *that* a program works but what components are most effective, for whom it is working best, and how it can be modified to serve a particular group.

Research designs for outcome evaluation

What designs for following the psychological model already exist in the literature and can be presumed to have worked?

Although random assignment is used occasionally, it is almost never recommended. Because people resent it so much, results may not be valid.

One resolution to this dilemma can be used by a program which offers several possible treatments, all of which might be good, and randomly assigns subjects among those treatments. This is very ethical, because you don't have anybody who isn't getting treated even though they are getting different kinds of treatment. The obvious risk of doing this is that all of your treatments are good. If so, you're going to end up with a case in which all of your people look exactly alike at the end, and you're not going to know whether you did anything for them or not, because you gave them a lot of services.

Kathryn Barnard's study of different nursing models is a very impressive example. The program involves three different models using nurses as intervenors with young parents; all three are very intensive approaches to providing intervention, and the results should come out that at the end of it they all look pretty much alike. I am convinced, however, that all of those three programs worked. The only thing I could suggest at this point is that you might look at something like your multi-problem group and see if maybe if you restricted it to certain kinds of subjects that certain programs work better than others. It's a risky strategy, but I wouldn't be using it if I didn't know of at least one case where it had worked well. Janet Hardy also used this model in the prenatal portion of her intervention. What she looked at was a very high quality regular prenatal program for high risk mothers at Johns Hopkins University Hospital and another high quality program which provided the same medical services and also provided social services such as career counseling, educational evaluation, and social help for adolescents having problems with their mothers—a much more personalized approach than you would get with regular high quality prenatal care. Reasoning that both programs were good, Hardy randomly assigned her girls to receive one of the two programs. There were medically significant different outcomes at delivery. The personal contact, over and above what was offered in good routine prenatal care, made a demonstrable difference to the girls.

In the Yale Child Welfare Research Program, Sally Provence and her group used what I would call the time lag design model. As part of a follow-up study of twelve year old children who had been in an intensive intervention program as infants and toddlers, we wanted to look at the people we would have served if the program still existed. Since the intervention group was selected by using medical record information, an identical approach was used to select as the control group persons who would have been the next people to be served had the project continued.

A similar design could be used by a program which had its approach planned but was awaiting funding. You could decide how you wanted to recruit clients, go ahead and recruit them and say "We don't have a project now, but we're hoping that we will have, and here's what it would involve. We couldn't start giving any actual services until a year from now, but are you willing to let us interview you, get some basic information about you, and then enroll you a year from now?"

What would you do in such a situation if you really think the major benefit is going to occur in the first year of life? When the money comes rolling in, you continue to recruit new subjects and treat them starting from the birth of their children. The first group of clients, of course, will only begin receiving services when their children are a year old. Comparing the two groups will give you some information about what indeed happens in the first year of life if no treatment is available.

While these evaluation approaches may be extremely worthwhile and much more appropriate for early intervention programs than the psychological model, I must emphasize that doing the research and gathering the data are as much work as doing the program and providing the services. If you're a service provider, how are you to find the time to do the research? And since whatever you love most is what you're going to do best, if you're in love with giving service, then you're going to do that and do it well, and let the research take a back seat when a crisis comes up. This is the right choice, of course, for a service provider. You really almost have to have somebody whose primary concern is the research, who will never let it slip; I don't see how that person can be a service provider.

The question of cost

I don't advocate what economists would call a cost-benefit analysis because I'm not sure I know well enough how to do it. What I do think is worth doing is calculating financial benefits that might be assocated with the outcomes of your program. Everybody can calculate how much a program costs: if you have outcomes with beneficial financial aspects, you may as well fight back and calculate those, as well.

We reached this conclusion quite by accident. In the course of the follow-up evaluation of the children in Sally Provence's Yale Child Welfare Project, the first thing I looked at was these children's IQ scores, and the very first thing I discovered was that there was absolutely no difference between the experimental and the control children years later. I went home depressed, but I knew that program did *something*. When I came back the next day and told my co-workers that I had found no difference in IQ whatsoever, one of them looked at me and said, "There are enormous differences in those children." She had gone in to examine children in school, not knowing who was experimental and who was control, but when we looked over the list, she said, "With this boy, I went in to pick him up to go down the hall to do the achievement testing, and the teacher begged me to take him out of the classroom: he drives her crazy. When I came back, she told me that he makes her life miserable, he's climbing the walls, he's 'hyper' and he's crazy. That kid is an absolute disaster." This is a very disturbed child, one who has been in special services all the time he's been in school. Another control child, we learn from his teachers, has gone on homebound instruction because he's been fighting teachers and he threatened to set fire to the school. And so on...

We decided to send another investigator to the schools, one who didn't know which kids were experimental and which were control, to try to find out how teachers felt about these twelve year olds and what services these children were needing from the schools. Now we found a whopping difference.

> *"If you have any outcome data that might legitimately affect health outcomes and particularly the use of expensive services like hospitalization or psychiatric care, document your findings, put a price tag on the services your clients didn't need."*

And it's not just that the experimental kids need *fewer* services; they also don't need the *expensive* services like court costs, pychological evaluations, and so on.

Whether this constitutes a cost-benefit analysis I don't know. I don't think we have any way of completely documenting all the costs and benefits that come out of intervention programs. But what you do have, you might as well use. If you have any outcome data that might legitimately affect health outcomes and particularly the use of expensive services like hospitalization or psychiatric care, document your findings, put a price tag on the services your clients didn't need. Among the families we serve, a small number tend to generate enormous costs which represent a large percentage of total health and 'service' expenditure in any community. If you can reduce the size or need for services of this costly minority, your intervention is probably cost effective.

The Infant-Parent Program at San Francisco General Hospital

by JEREE PAWL

Since I will be describing a process of developing an evaluation component for a specific service program, let me tell you something about the nature of that effort.

The Infant-Parent Program at San Francisco General Hospital is an effort of the University of California/San Francisco, Department of Psychiatry. Begun in 1977, the program is a direct outgrowth of a research-demonstration program that existed at the University of Michigan from 1973 through 1979, directed by Selma Fraiberg. Unlike the Michigan program ours was not supported through research or demonstration grants but was instead funded primarily as a service and education program by a coalition of City, University and Foundation interests. It was intended to serve a wide range of families and agencies and was designed to be flexibly responsive to the many different kinds of requests that we received for our services. This allowed for a kind of free-market effect, which in turn permitted the program to be shaped both by the requests and referrals coming to it and by its ability to respond satisfactorily to those requests. As a result, the Infant-Parent Program currently exists as a resource for assessment, treatment, education and consultation (both case centered and programmatic) to a whole range of agencies throughout the City which are concerned with infants and toddlers up to 3 years. This group includes all the various branches of the Department of Social Services (such as Children's Emergency Services, Protective Services, reunification and adoptions) as well as the Juvenile Court, public health nurses, private pediatricians and psychiatrists, numerous hospitals, and a variety of community mental health and social service agencies. The program is located at the City/County hospital, San Francisco General, and the primarily low-income, multi-ethnic population of that hospital is clearly reflected in our caseload.

From the beginning of the program we have tended to attract the desperate—both in terms of the state of the referring source and of the families referred. Although we would have hoped to serve a very wide range of families, those in a position to refer to us throughout the hospital and the community tend to refer those cases which are keeping them awake at night. They have done so with some apologies but with no subsequent changes in their behavior. As a result, we have received referrals primarily of multi-problem families who are awash in socioeconomic problems and substance abuse and whose infants and toddlers are often either gravely at risk for severe abuse and neglect or are already victims. These are people whose lives and often character structures preclude at least initially the organization necessary to keep regular office appointments and who need consistent and unswerving outreach efforts to establish any connection at all. As we are a home visiting program, these efforts are possible for us as they are not for other Community Mental Health outpatient services, where all visits are office based. But this has made our fit into the Community Mental Health system awkward and difficult, though not impossible.

The demand for an overall evaluation component

Initially, our sources of funding were interested in outcome primarily in terms of the numbers of families that we served, the ethnicity of those families, our availability to other agencies, the source of our referrals and the numbers of workshops and presentations that we gave.

When we became a bona fide Community Mental Health outpatient agency, however, we needed to add an overall evaluation component to our existing internal and more informal evaluation process. Community Mental Health, and later the Children's Trust Fund, from whom we also receive money, wanted to know something about how "successful" we were. They also continued to want demographic and descriptive data, but this was no longer sufficient. They were not content with our "clinical impressions", or with a report on a group of 42 failure-to-thrive patients that we had studied, or with individual case studies—though all of this would continue to interest them. It should be noted that no additional money for evaluation or research was forthcoming; our expanded evaluation component, therefore, could not be too costly in terms of time.

In short, the new process of evaluation had to be integrated into the work in some sane way. Time constraints alone dictated this. The evaluation could not involve a significant additional effort beyond the effort to construct and institute it. Nevertheless, we had to find a way to give our funding sources the numbers they wanted and to make these numbers meaningful both in terms of the amount of service delivered and the successfulness of the intervention. The evaluation also needed to be truly reflective of what we did.

With all families our work includes some mix of case management, procuring of a range of services in a benign way, concrete and emotional support, developmental guidance and infant-parent psychotherapy. Obviously, we did not make the same combination of efforts with every family and what might constitute an intervention success with one family might be a given with another. Thus, the evaluation would need to be individualized yet organized in a way that could meaningfully describe the general effectiveness of our program. This meant an articulation and ongoing tracking of what it was we were really trying to accomplish with each family referred.

> *We had to find a way to give our funding sources the numbers they wanted and to make these numbers meaningful both in terms of the amount of service delivered and the successfulness of the intervention. The evaluation also needed to be truly reflective of what we did.*

Our evaluation procedures: Goals and objectives for assessment and treatment

From these demands for information and these constraints on our resources, our rather rudimentary, but very useful and quite satisfying, evaluation procedures have developed. We had to sit down, think about what we did, think about our families, realize that each one was unique, and then develop some system of describing what our intent was and how we were proceeding in terms of meeting that intent.

For purposes of supervision and education of our trainees as well as writing case studies, we continued to use detailed process notes—but for purposes of evaluation these were far too cumbersome. The time needed to transform detailed and diffuse material into a condensed form permitting ready evaluation is too great. This is true even for sub-groups of cases, or for tracking specific issues in certain kinds of cases (for example, the development of transference). Such projects will always require extra funding or the specific interest of some staff member who can somehow devote the time to planning and executing the study.

Currently, our evaluation charts describe each family and detail the reason for the referral. All of our contacts with families begin with an assessment which includes five to six home visits over an approximately six-week period, along with a videotaped formal developmental assessment of the infant or toddler with an additional half-hour free-play session. At the beginning of the assessment, the objectives for that assessment are outlined based on the referral. Some additional objectives may be added during the assessment period. At times this assessment will comprise our only involvement with the case. We are frequently asked for an evaluation which includes a written report with recommendations, followed by court testimony, for example. Parenthetically it may be of interest that we are sometimes asked to evaluate the same child in a variety of situations such as with the biological parent, with the foster parent and in a nursery school or daycare setting.

Whatever the task, we always outline clearly what we are being asked to do and what we will need to do in order to accomplish that. This is part of the process of delineating the specific objectives for this particular undertaking. All of the objectives derive in some sense from general capacities in which the particular family is seriously deficient. Sometimes an objective may be extremely rudimentary indeed. For example, many of our families are extraordinarily difficult to engage, and it may be an objective to establish a working relationship with the family such that an assessment may take place. Absurd as this may sound as an objective, it is an outcome in which both we and our funders are clearly and justifiably interested.

We are also interested in more detailed information about the circumstances associated with success and failure in meeting our objectives. For example, we were successful in establishing a relationship in 92% of the cases for which an evaluation was requested. We were successful 82% of the time in engaging a family where ongoing treatment was the stated goal of the referral source. We were successful in only 78% of the cases when the referral was court mandated and the Department of Social Services was involved. We were successful in completing the assessment 100% of the time if two visits were completed. Such details are instructive not only to our funders but to us. We have learned, for example, that we cannot be shy about addressing families' negative feelings about a court-mandated assessment and that we will lose the family unless every opportunity to support their anger and dismay is fully exploited. We certainly knew this, but there is knowing and knowing. When we learned in which particular instances we were being less successful, we renewed our efforts to find out why. Bu studying the detailed process notes and the evaluation chart narrative, we learned that our successful efforts included a heavy emphasis on interpreting the negative feelings of the parent; in unsuccessful cases, such interpretations were less frequent. This review helped us to realize, also, that increasingly skillful intervention by the therapist could and should be as much of an objective as improvement in a child, family or relationship. The review pointed out things we needed to teach and to emphasize in our supervision of beginning trainees.

We had to sit down, think about what we did, think about our families, realize that each one was unique, and then develop some system of describing what our intent was and how we were proceeding in terms of meeting that intent.

Another goal which might be typical for us would be to achieve a good understanding of the functioning of the parent as a person and as a parent. As far as the parent as a person is concerned, we organize material from all sources at the end of the 6 week assessment period so that we can, ideally, describe his or her defensive functioning, object relations, affective functioning, and cognitive style and give a preliminary psychiatric diagnosis. We will also have organized our information about each parent's functioning as a parent into various standard descriptive categories. This description covers both physical care and emotional responsivenss. For purposes of *evaluation* we will either have failed to achieve an understanding of the parent's functioning, will have made progess towards this, or will have successfully achieved this objective during the assessment.

For each family, we have a number of objectives, though not always the same number, and of course we have a number of families. As a result, we are able to say very simply but very clearly how successful we were in achieving the objectives and goals of our assessment both for any particular family and for the cases overall. In our latest report to the Children's Trust Fund, for example, we could report that we had been successful 83% of the time in meeting our objectives for assessments. We are also able to describe the areas or kinds of objectives we are consistently more or less successful in achieving.

As we move into the evaluation of treatment, essentially the same procedures are used, though the objectives need to be of a different kind and at different levels in order to be most useful. Initially, we tried to limit our objectives to a single level of discourse, but we found this both too inflexible and not nearly as useful as our current approach. A list of objectives for a particular family at the beginning of treatment might include: 1) child will be in daycare; 2) mother will stop hitting child; 3) mother will become more aware of the child's own experience; 4) mother will become a member of the mother's group at a particular center; 5) the child's range of affect will be increased; 6) the child will show less wariness around his mother; 7) the mother will show some signs of experiencing connections between her own history and her current experience with the child. Such a list exists because it has evolved from what one has seen and documented in the process of the assessment. Based upon those observations, these objectives are determined to constitute helpful and meaningful achievements for this particular family. They are general objectives in that they are relevant to the entire population, but for this specific family they represent the remediation of significant deficiencies. Although the various levels of these objectives require different kinds of evidence and observation in order to support either failure or success, all the objectives warrant inclusion.

The documentation necessary is recorded weekly in a brief description of each visit which is specifically organized around and addresses each objective. This note records the observations of information relevant to each objective. In addition, there is, as I said, a detailed process note of the visit, and that can provide useful back-up documentation though it is *not* part of the evaluation chart. Finally, we have an evaluation tracking *form* where the objectives are listed and where check marks are placed under either success, progress or failure. Over time, objectives are added as they become relevent. There is a simple dating procedure to record when each objective is formulated and when it is evaluated. All objectives are addressed individually in the narrative form every three months. This allows progress to be tracked and new objectives to be added.

The long range goals for each family are more comprehensive and are, in a sense, made up of the specific objectives. That is, the objectives must be achieved in large measure before the overarching goals could be met. For treatment, the goals are rated at the time of termination. And at that time, rather than merely checking failure, progress or success, we use a rating scale to indicate a judgment about the degree of success, if any. The goals are individualized for each family and could actually be rated by anyone using the evaluation chart. Two raters could achieve reasonably good agreement about whether an objective has been met simply by utilizing the narratives which address the objectives as these include relevant material from all sources. They could also agree regarding the level of success in terms of the goals which have been achieved.

Preserving clinical complexity

It is important to note one other important factor. The same objective may appear in several charts but have different meanings depending upon the case. "Getting a child into daycare" may be an objective for two families. For one family, this objective may necessitate a new level of organization and prioritizing which will require a significant amount of support to accomplish. For another family, getting a child into daycare may involve dealing with severe separation problems. This may make getting this child into daycare extremely arduous and an extraordinary triumph. Formulating daycare as an overarching goal rather than a specific objective might be one way to reflect this. The difference is likely to be better highlighted by the additional objectives, though, as well as by a more encompassing goal. For the latter family, for example, the comfortable separation of the mother and child may be an overarching goal. Getting the child into daycare may be the external manifestation of an enormous improvement in the mother-child interaction and of the mother's ability to experience her child as a separate and autonomous person. This same kind of change may be indicated by such objectives as "mother no longer sleeps and bathes with her toddler".

Since the actual meaning of the objectives and goals will be evident from the narrative and from the range of objectives, our evaluation system preserves clinical complexity while maintaining a simplicity that allows objectives and goals to be listed. It is in the interest of maintaining this richness and yet achieving clartity that we have a mix of levels of objectives. This mix does not prove a problem because each objective rests on the literally observable or on what one could readily conclude and agree about in a consensually validated way. Each is thus fairly readily measured. This mix allows us to retain the psychological complexity and yet build a foundation upon which agreement through rather simple counting or observation or listing can be based. This has seemed to us very important because we wanted to avoid reducing what we included to the most simplistic level. Also, to do that would be a bit like "teaching to the test".

Since the actual meaning of the objectives and goals will be evident from the narrative and the range of objectives, our evaluation system preserves clinical complexity while maintaining a simplicity that allows objectives and goals to be listed. It is in the interest of maintaining this richness and yet achieving clarity that we have a mix of levels of objectives.

Effects of the evaluation procedure on clinical work

It has been our experience that having these clear objectives to which one must address oneself weekly in a narrative report has a very salubrious effect on the clinical work. It continually reminds one of the fact that the process is not only a process but also has distinct objectives. This evaluation procedure helps to keep this in focus. It sometimes also suggests connections to be exploited in the material which might otherwise be overlooked.

At the same time the evaluation procedure does pose a threat. If "success" is too narrowly defined as that which is represented by a single objective or even goal then its achievement could be pursued at any cost--losing the point entirely. The *means* by which one achieves the objective is in one sense the whole point of the intervention: therefore it is important that some of these *means* also be included as objectives. That is another reason why there should be a healthy mix of levels in one's objectives. For example, if one objective is to decrease harsh, overcontrolling behavior on the part of a mother it is necessary to place by its side, for instance, the objective to increase the sensitivity to the experience of the child. This will ensure that you don't steer too insensitively toward one objective at the expense of another. The evidence for both objectives can be quite concrete. "Increased sensitivity to the experience of the child" can appear clearly in the narrative though the signs of it cannot be specified in advance. Thus, we could not have as an objective that a mother will suddenly state that her toddler feels very left out when she feeds the baby. There exists such a range of possible attributions and circumstances that this particular one could not be detailed ahead of time; yet it will clearly fit under "increased sensitivity to the experience of the child".

Whatever evaluation method you devise, if it teaches you nothing and fails to shape your efforts, it is probably not properly reflective of what you actually do nor what you hope to achieve. It really needs to be an organic expression of the work.

I should state to you at this point that I am acutely aware of the fact that much of what we wrestle with does not bear a one-to-one relationship with the circumstances and situations in which many practitioners who see infants, toddlers and their families generally work. It is probably not such a triumph for many professionals to achieve regularity of meetings or an alliance with parents as it is with us. Nonetheless, I think our point of view about evaluation is relevant for several reasons. We believe that this kind of evaluation in general should be structured from the inside out—growing out of an understanding of the specific functioning of the family of concern. In addition, where the functioning of an infant or child is involved there must be a recognition of the importance of the centrality of a "relationship" and of the idiosyncrasy of each dyad. Those few handicapped infants whom we come to see are, most typically, though not always, living in situiations of some neglect and abuse. Our goals in these cases do not differ in general from what we hope to achieve with the other infants and toddlers and parents we see. All the handicapped infants and their parents in our program are simultaneously enrolled in infant programs responsive to their disabilities (though most often this is only as a result of our efforts).

It is worth emphasizing however, that even with these disabled infants our focus has been the developing relationship. In that regard--in terms of the importance of the quality of the interaction between parent and child--there is no difference between the physically able and disabled children and their parents. As some convincing evidence of this, I should offer the following piece of information. Recently one of our trainees finishing her year in our program was hired by an excellent San Francisco program specializing in work with multiply handicapped infants and their parents. She was hired specifically because of her expertise in infant-parent interaction in order to add a new dimension to their program's ongoing work. She has, to some extent, refocused the program to include very careful observation of the parent-infant interaction, work with infants and parents together, and a complementary de-emphasis on the more exclusive focus on a curriculum for the baby. There is a clearer emphasis on elaborating skill-building in the baby in the context of the natural and enhanced interaction of the mother and baby and on improving the mother's observational skills, her satisfaction in her baby, and her own sense of competence. When the quality of the relationship between parent and child becomes a major focus in such a program, then many of the things I have been suggesting as objectives for our families would fit these families equally well. Certainly they do in those families with disabled infants within our own population.

Conclusion

I have attempted to describe something of the process we went through in response to a need to demonstrate to our funding sources the efficacy of our efforts with our families. We found the process not only satisfying to them but helpful to us, and that may be the final criterion. Whatever evaluation method you devise, if it teaches you nothing and fails to shape your efforts, it is probably not properly reflective of what you actually do or what you hope to achieve. It really needs to be an organic expression of the work, clumsy and inelegant as it may seem. It is not statistically sophisticated nor does it in any way qualify as "outcome research", but most of us do not have the funding to do anything that would resemble that. Nonetheless we can be proud of an evaluation process which truly expressed the functioning and purposes of the work of a program. Ideally, the evaluation component needs to reflect *first* one's own questions and only second those of one's supporters. I can assure you that if it does the first it will also do the second.

Discussion

I. Community and state issues: organizing for better program evaluation

Meeting participants included local and county-level service providers and state agency program administrators from Hawaii, Ohio, Texas and Washington. Their experiences and concerns reflected a number of important issues involved in public programming for disabled and at-risk young children and their families.

- *Comprehensive intervention is costly - how can long-term benefits be justified?*

As long ago as the early 1970's, professionals in **Hawaii's** Infant and Child Development Programs were recognizing that therapeutic intervention for young children with developmental problems would not have long-lasting effects without changes in parental knowledge, skills and attitudes about caring for these children. Although many experts in the field agree with these clinically based conclusions, the dearth of outcome data from comprehensive intervention efforts with this population has raised questions about the value of multidisciplinary approaches to families of infants with disabilities, approaches which are costly in the short run although they are likely to be ultimately cost-effective. Selected testing of a family-oriented service model before adopting it as standard practice may be a useful approach. States can also encourage diverse early intervention approaches for specific populations and measure their relative effectiveness. This evaluation model presents no ethical dilemmas if all clients receive services which offer promise of effectiveness. In five year planning for services to disabled and at-risk infants, toddlers and their families, evaluation needs to be raised as a major issue.

- *Offering and soliciting community feedback*

As staff of two new local early intervention programs in Ohio described their services, the importance of feedback to other providers in the comunity as well as to program staff became clear. Since early intervention research has generally not followed the medical model of controlled clinical trials to which most physicians are accustomed, service providers working with pediatricians and other physicians need to design evaluation measures which will be meaningful both to themselves and to the medical community. Numbers - of appointments kept or referrals used, for example, - are convincing. While interviews with parents may be a good way of assessing changes in parental competence, asking for evaluation of changes from personnel from other agencies, friends of the parent, or the mothers of adolescent parents may also provide compelling documentation of changes.

- *Developing a statewide research agenda; elaborate and modest models*

With a legislative mandate to devise outcome measures for its new interagency early intervention program, the state of **Texas** is involving many organizations and institutions in developing an evaluation proposal. Maternal and Child health personnel at the regional level, grantees of the Handicapped Children's Early Education Program, high-risk infant follow-up programs, directors of the four agencies collaborating in Texas' Early Childhood Intervention Program, state universities, and the University Affiliated Facility are all involved.

Collaboration at the planning stage may make possible common reporting and aggregation of data from a number of programs, a valuable process which is difficult, if not impossible, to do retrospectively. Instruments measuring key indicators of child, family and health status need to be devised ahead of time for common reporting. The questions then are raised "Who owns the data? Who has access to it?"

Still, as **Ohio** participants pointed out, a program must start with its own evaluation agenda, not what the State wants reported. Too many programs are simply monitoring rather than evaluating themselves.

Even with no state funds allocated for program evaluation, **Washington** State has found it valuable to survey data currently being collected by individual programs in order to plan future intervention and evaluation activities which may be carried out on an interagency basis.

Ohio is also trying to work out some common reporting data aggregating key indicators for evaluation purposes, concerned with questions of who owns the data and how can it best be shared, while respecting confidentiality.

This is a particular problem when the data include family assessment measures. The state has been asked to compare "the effectiveness" of various county programs, based on their relative scores on common reporting forms, even before the forms have been well tested and validated. States also have the option of comparing demographic outcomes in counties with rich services to those in counties of similar populations with very few services. Longitudinal studies are extremely important in documenting program effectiveness and should be supported by states.

II. Generic Issues

The following issues are of concern to all those working in early intervention, whether service providers, researchers and program administrators or those setting policy and allocating resources for disabled and at-risk young children and their families. Knowledge gained from program evaluation efforts should guide and improve individual programs and also make an impact on the design and structure of our country's services for this population.

• *The need for providers and administrators to broaden the range of outcomes used, and to educate others as to their importance*

Clinicians and program directors need to educate the public and policy makers about what "success" means in early intervention, and the questions to ask in order to evaluate program effectiveness. It is important for policy makers to understand that "no change" can be a good outcome in many situations. While early intervention programs may help some infants and families make dramatic gains in functioning and/or avoid impaired development, *preventing decline* in the child or disintegration in the family can represent an equally major achievement in other instances. Similarly, short-term positive outcomes of early intervention programs are valuable. Service providers will never be able to offer babies "vaccinations" against future adversity, but positive long-term outcomes in the pre-school years, in later childhood and in adult life are more likely when the infant and family can build on a series of short-term successes in the earliest years of life.

• *The myth of an ideal single early-intervention program for all children*

As researchers come to understand more about the many biological, temperamental, familial, socio-economic and environmental factors which play powerful and complex roles in influencing an infant's development, they are confirming the experience of clinicians that no one program can ever be appropriate for all infants or all families. The stresses children and families face and the unique blend of strengths and vulnerabilities with which they face these stresses are simply too varied.

Since we do know that early intervention approaches are *generally* effective, research and evaluation efforts now need to address the more specific questions of which programs work best for whom and why. Finding the answers to such questions will eventually allow much more appropriate targeting of services to those who need them, but diversity should be valued rather than deplored. The richness of intervention programs (transdisciplinary team vs. a single, less highly trained intevenor) also need comparison evaluations, looking at their work with various kinds of clients.

• *Clinical impressions as an adequate basis for quantifiable data: "Take it out and shine it up"*

Clinical impressions and quantifiable data can both be useful in evaluating the effectiveness of early intervention programs; they need not involve intrusive interviews or potentially threatening questionnaires. The satisfaction of intervenors with their own efforts, observable behavioral changes in the way parents interact with their own children, and providers' assessments of clients are all good indicators of program effectiveness. Quantifiable data can include measures as diverse as rates of kept appointments, changes in related physical problems or emergency room use, length of time between pregnancies, the amount of parental involvement with the child, the quality of the relationship between parent and provider, or the ability of the family to negotiate the social service system independently.

Providers often notice changes in families during the intervention process which can be unorthodox, yet valid, measures. For example, how has behavior of the oldest female sibling of the handicapped infant (who often bears a considerable burden of the stress associated with a younger child's disability) changed since the beginning of the intervention program? What does this change say about program impact?

For this kind of monitoring of program quality, outside evaluators are not really needed. Rather, program staff need to choose and articulate their intervention objectives and determine how to measure their degree of success in concrete, objective, observable ways. Often this is a process of operationalizing formally the informal review and monitoring of practice already ongoing in an agency -- of "taking it (this process) out and shining it up."

• *Unserved populations: a difficult area for evaluation*

Examining the disabled and at-risk infants, toddlers and their families who are not being well served or even reached at all by existing programs is a difficult task but one that is crucial for statewide planning and allocation of resources. Such assessment may be undertaken more easily by state agencies or interagency groups than by local programs, but should be reflected in planning and evaluation at all levels.

Planners may want to pay particular attention to families with few sources of social or emotional support. While these families may be hard to find and engage, service programs may make a dramatic difference in their lives. Currently, fathers seem frequently left out of both service programs and evaluations, despite their contributions to a family's coping capacity.

- *The problem of cost benefit analyses*

Meeting participants agreed that convincing cost-benefit analyses of early intervention services are probably not possible at this time. The far-reaching and long-lasting impacts of untreated disability or impairment on individual, family and society are as incalculable as are the ultimate effects of the complex variety of services now available. It should also be recognized by state administrators that provision of needed services to vulnerable populations has to do with basic humanitarian issues. The question is how to do the job more effectively, not whether to do it at all.

- *The real world: natural experiments and promising hypotheses*

In planning and conducting program evaluations, it is useful to look for situations which represent "natural experiments." For example, the circumstances of a population which should have received services in the past but did not get them can be compared to outcomes in a population now served by newly available programs.

While it may be impossible to identify precisely the ingredients in a program which contributed to its success, clinicians may generate hypotheses about such questions for further investigation in highly controlled experimental studies by outside evaluators.

- *Research and evaluation: everyone's responsibility*

Meeting participants agreed that every early intervention program has the responsibility to undertake *some* evaluation system to provide guidance and direction to staff and program planners. National outcomes evaluation efforts should not be relied on for individual program justification. Intake assessments, daily logs, process notes, case conferences and similar elements already exist in many service programs. The data they provide can be used to evaluate services if providers take the time to articulate and use this information. Evaluation efforts which involve staff, recognize program accomplishments and point out where improvements in service can be made can improve both program practice and staff morale itself, another key factor in successful intervention.

Although commercial enterprises routinely allocate 10-20% of their budget for research and evaluation, early intervention programs seldom specify evaluation as a separate budget item. In addition to the process evaluations service providers can do, longitudinal, carefully controlled studies should be carried out by outside researchers backed by a major commitment of funds. National efforts are also needed to aggregate rich but specialized data from individual programs.

References

Dunst, Carl J. and Rheingrover, Regina M. — "An analysis of the efficacy of infant intervention programs with organically handicapped children" Evaluation and Program Planning, Volume 4:287-323, 1981.

Dunst, Carl J., Snyder S. and Mankinen, M. — Efficacy of early intervention, in M. Wang. H. Walberg and M. Reynolds, editors *Handbook of Special Education: Research and Practice* (Oxford, England: Pergamon Press, in press)

Meisel, Joan — *Programming for Atypical Infants and Their Families: A Guide to Program Evaluation* (New York: United Cerebral Palsy Association, Inc. 66 East 34th Street, New York, NY 10016, 1977)

Office for Maternal and Child Health, DHHS — *Guidelines for Early Intervention Programs* (Washington, D.C., 1980)

Seitz, Victoria, Rosenbaum, Laurie K. and Apfel, Nancy H. — Effects of family support intervention: A ten-year follow-up *Child Development*, V.53: 376-391, 1985.

NATIONAL CENTER FOR CLINICAL INFANT PROGRAMS

733 Fifteenth Street, N.W., Suite 912
Washington, D.C. 20005
(202) 347-0308