Evidence-based Policy:

Strategies for Improving Outcomes and Accountability

North Carolina
Family Impact
SEMINAR

Center for Child and Family Policy, Duke University

Evidence-based Policy: Strategies for Improving Outcomes and Accountability
A Briefing Report prepared for the North Carolina Family Impact Seminar

North Carolina Family Impact Seminar briefing reports and other Center for Child and Family Policy publications are accessible on the Center's Web site: www.childandfamilypolicy.duke.edu.

Evidence-based Policy:
Strategies for Improving Outcomes and Accountability

Table of Contents

Statement of Purpose and Focus

The **purpose** of the North Carolina Family Impact Seminar (NCFIS) is to provide objective, nonpartisan, solution-based research on a topic of current concern to state leaders. The seminars purposefully address how policies and practices impact children and families. Legislators and legislative staff guide the topic selection each year, based on their own concerns and those of their colleagues and constituents, as well as their knowledge of what is likely to be addressed during that year's legislative session.

NCFIS includes annual seminars, briefing reports and follow-up activities designed specifically for state-level policymakers, including legislators and legislative staff, the governor and executive branch staff, and state agency representatives.

Through NCFIS, research, information and insight related to policy, practice and programs are presented via two avenues:

- Experts who present and interact with stakeholders at the seminar, and
- A briefing report produced specifically for the seminar.

NCFIS opens the door for ongoing exchanges between legislators and the experts who speak at the seminars; researchers, faculty and staff of Duke University's Center for Child and Family Policy, which convenes the state's seminars; and a broad range of stakeholders concerned about the issue, including members of the executive branch, directors of state and local government agencies, leaders of nonprofit agencies, and researchers and scholars from Duke University and other institutions of higher education. The briefing report is disseminated to this broad audience and is made available on the Center's Web site.

Family Impact

Family Impact Seminars encourage policymakers to consider the impacts of policies on families in the same way that they routinely contemplate the economic and environmental impacts of policies.

The first step in developing family-friendly policies is to ask the right questions:[1]

- What can government and community institutions do to enhance the capacity of families to help themselves?

- What effect does (or will) this policy (or program) have on families? Will it help or hurt, strengthen or weaken family life?

These questions sound simple, but they can be difficult to answer.

The 2009 seminar **focus**—evidence-based policy— is a timely and critical topic for several reasons:

- North Carolina, like many states and the nation as a whole, faces a severe budget challenge. This makes it more important than ever to ensure that funding and other resources are targeted to programs "that work."

- Legislators increasingly point to the need for evidence-based policies and programs, yet there is not a universally sound understanding among policymakers or program advocates of what "evidence-based" means in practice.

- The North Carolina General Assembly established a Program Evaluation Division in 2007 highlighting the growing interest in and expectations regarding outcomes and accountability.

Family Impact Questions Regarding Evidence-based Policy

- What strategies can the North Carolina General Assembly adopt to enact evidence-based policy routinely?
- How would an enhanced commitment to evidence-based policy lead to wiser investment of public funds and better outcomes for citizens?
- How does the increased attention to outcomes and accountability intersect with the growing demand for evidence-based policies and programs?
- Are there components of an evidence-based policy approach that are particularly well suited to North Carolina?
- Should North Carolina policymakers require programs they fund to demonstrate a common set of characteristics that lend themselves to the implementation of evidence-based interventions?

_____

[1] Policymakers and other stakeholders may wish to consult *A Checklist for Assessing the Impact of Policies on Families*, developed for the Policy Institute for Family Impact Seminars and available at www.familyimpactseminars.org/fi_checklist_aipf.pdf. The Center for Child and Family Policy does not endorse this checklist, but offers it as a resource for policymakers.

Executive Summary

Every day, policymakers make decisions based on information from diverse sources. Using an evidence-based policy approach can improve the decision-making process. Evidence-based policy relies on and uses high-quality research findings. Furthermore, by attending to program implementation and evaluation, legislators can apply evidence-based methods to craft policy that is more likely to yield effective programs and practices.

Across policy areas, the use of strong evidence to choose interventions, prioritize funding and implement programs will enhance the likelihood of positive outcomes. Given limited funding resources, strategic support of proven programs is all the more critical to maximizing the programs' benefits. Cost-benefit studies have identified programs in many areas of policy that, when well implemented, can achieve benefits that significantly outweigh costs (with some examples showing net *benefits* of up to $30,000 for every dollar spent on a participant). However, all policy areas also have programs that are not cost-effective (with some examples showing net *costs* of up to $50,000 per participant).[1] Careful evaluation and review of the evidence for program effectiveness are critical to fiscal and social outcomes.

This briefing report provides an introduction to the rationale, criteria and strategies for evidence-based policy. It also provides concrete examples and lessons learned from successful evidence-based policy initiatives in North Carolina and Pennsylvania. This report consists of five briefs:

- Brief 1 introduces evidence-based policy, including the rationale for using rigorous evidence, strategies for identifying programs that work and procedures for building evidence for newly developed or untested practices.

- Brief 2 focuses on pilot programs, often a key aspect of evidence building. The brief presents 10 questions for policymakers to consider in their efforts to design and fund pilot programs that are capable of generating valid information about program effectiveness.

- Brief 3 presents a concrete example of evidence-based policymaking in North Carolina: the development and evaluation of the Graduated Driver Licensing system.

- Brief 4 discusses the importance of implementation in achieving optimal outcomes with evidence-based programs. The brief also presents the work of a unique collaborative of stakeholders working to support high-quality infrastructure and program dissemination in North Carolina.

- Brief 5 describes the principles and strategies behind Pennsylvania's statewide effort to increase reliance on evidence-based programs and presents data on resulting cost savings.

Enhancing and complementing the information and strategies highlighted in the briefs, the report also includes the following items related to evidence-based policy:

- A list of acronyms,
- A glossary, and
- A list of additional resources.

As this briefing report illustrates, increasing the use of evidence in policymaking yields both monetary and "family impact" benefits. While the required investments to "do" evidence-based policy in the short term may be significant, the long-term benefits of doing so will almost certainly be more significant.

---

[1] Aos, S., Lieb, R., Mayfield, J., Miller, M., & Pennucci, A. (2004). *Benefits and costs of prevention and early intervention programs for youth*. Olympia, WA: Washington State Institute for Public Policy.

4

Using Rigorous Evidence to Improve Government Effectiveness:
An Introduction

Katie Rosanbalm

This introduction to evidence-based policy is for policymakers, agency officials and program administrators. In this brief, the reader will learn the following:

- The rationale for using rigorous evidence to inform decision making and policy development;
- Strategies for identifying evidence-based programs; and
- How to build an evidence base for newly developed or untested practices.

While the context for this brief is social policy, the principals for evidence-based policy cut across numerous policy domains.

What counts as evidence?

Evidence-based policy is public policy informed by rigorously established objective evidence. The goal of evidence-based policy is not simply to increase reliance on research results to inform decision making, but to increase reliance on "good" (i.e., rigorous) research. The first step in using evidence-based policy is learning how to objectively weigh information to determine its value as evidence.

The plural of anecdote is not data.

Stories (from neighbors, friends, family, the media, constituents, etc.) often provide strong messages about the positive or negative effects of various interventions and programs. Program advocates may describe individuals whose lives improved dramatically after participating in a particular program, and it is tempting to replicate the program to bring these benefits to others. But do these anecdotes and case studies provide definitive evidence of program effectiveness? Do they provide sufficient data to support program dissemination? In a word, "no."

For a program to earn the classification "evidence-based," it must have been rigorously tested and found to achieve its stated outcomes effectively. While untested programs *may* result in positive outcomes, without rigorous research it is not certain whether they do or not. Equally important, it is not certain what types of people or populations the programs benefit.

Top-tier evidence-based programs are those proven in well-designed and well-implemented randomized controlled trials, preferably conducted in natural community settings, to produce sizeable, sustained benefits to participants and/or society. Ideally, similar positive findings of such programs will have been observed by more than one evaluator and in more than one community. For the purposes of replication, programs also need, at a minimum:

- Clear written guidelines for implementation (i.e., a manual or curriculum), and
- Mechanisms for monitoring intervention fidelity.

Though not considered top-tier, programs backed by less strong evidence may be highly promising and worth pursuing with rigorous evaluation to verify whether they would maintain their value if brought to scale.

Given limited funding resources, strategic support of proven programs with solid evidence will maximize spending effectiveness. Every dollar spent on an ineffective program is a dollar that could have been spent on an effective one. This is not a call to stop developing and funding new and innovative programs. However, new programs are most likely to succeed if they are informed by past successful efforts and include careful piloting and rigorous evaluation prior to wide dissemination.

Why does evidence-based policy matter?

There are many good ideas, many intervention models and many skilled individuals with the best of intentions to provide services to help individuals improve their lives. Yet throughout the nation there has been little progress in key areas of policy over the past several decades. Consider the following:

- Government data on long-term trends in K-12 education show limited progress in raising reading, math and science achievement over the past 30 years;
- The US poverty rate today is higher than it was in 1973; and
- Despite some recent improvements, government data show limited overall progress in drug or alcohol abuse prevention and treatment since 1990.

With all of the research activity and intervention development that has occurred, there has not been substantial change in these and other areas of policy and service provision. Evidence-based policy, however, holds a key to positive change.

Evidence-based policy provides an effective mechanism to establish, in a scientifically valid way, what works or does not work, and for whom it works or does not work. With this structured approach to evaluation, knowledge can be used to improve practice, allowing successful programs to develop iteratively over time. Without this approach, interventions go in and out of practice, little is learned about what works, and the effectiveness of social programs does not advance significantly over time. Rigorous evaluation can end the spinning of wheels and bring rapid progress to social policy as it has to the field of medicine.

Rigorous evaluation has identified some highly effective interventions with returns, both financial and individual, far surpassing the investment. Examples include the following successful programs:

- The Nurse-Family Partnership (nurse home visitation for low-income, pregnant women) produced 40-70 percent reductions in child abuse/neglect and criminal arrests of children by age 15.[1,2]

- The Riverside GAIN Program (to move welfare recipients quickly into the workforce through short-term job search and training) increased single-parent employment and earnings by 40 percent at five-year follow-up.[3]

Cost-benefit studies have identified programs in many areas of policy that, when well-implemented, can achieve significantly more benefits than costs (with net benefits of up to $30,000 for every dollar spent on a participant).[4] However, in each of these policy areas there are many programs that are not cost-effective (with net *costs* of up to $50,000 per participant). Careful selection of interventions is clearly critical to fiscal and social outcomes.

Conventional wisdom is overturned.

Policymakers and other stakeholders can learn much from medical research, which has shown that conventional wisdom about "what works" is often wrong. Following rigorous evaluation, ineffective interventions have been modified or halted, paving the way for ongoing development of new treatments that can be proven effective. For example, well-implemented randomized controlled trials (RCTs) have shown that the medical interventions listed in the first table below and believed effective for decades are in fact ineffective or harmful. Similarly, rigorous studies of social programs have found that many popular interventions have weak effects, no effect or even adverse effects. (See second table below.)

Conventional Medical Treatments Found Ineffective by Randomized Controlled Trials	
Intervention	**Negative Outcome(s) Revealed by RCT**
Intensive efforts to lower blood sugar of type 2 diabetics to normal levels in order to prevent heart disease	Depending on the method used to lower blood sugar, either ineffective or harmful (increased risk of death)[5]
Hormone replacement therapy for postmenopausal women	Increased risk of stroke and heart disease for many women[6,7]
Dietary fiber to prevent colon cancer	Ineffective[8]
Stents to open clogged arteries	No better than drugs for most heart patients[9]
Beta-carotene and Vitamin E supplements ("anti-oxidants") to prevent cancer	Ineffective or harmful[10]
Oxygen-rich environments for premature infants	Increased risk of blindness[11]
Promising AIDS vaccines	Doubled risk of AIDS infection[12]
Bone marrow transplants for women with advanced breast cancer	Ineffective[13]

Popular Social Programs Found Ineffective by Randomized Controlled Trials	
Intervention	**Outcome(s) in RCT**
Vouchers for disadvantaged workers, to subsidize their employment	Large negative effects on employment rates, likely due to stigma caused by the methodology of supplying the vouchers[14]
Scared Straight, a program to prevent juvenile delinquency	Small *increase* in subsequent criminal activity by participating youth[15]
Drug Abuse Resistance Education (DARE)	Ineffective in preventing substance use (now being redesigned)[16]
Even Start family literacy program	Child and parent changes in literacy equivalent to those of a control group[17]
New York City vouchers for disadvantaged youth (K-4) to attend private school	Weak or no effects on student achievement[18]
Job Corps academic and vocational training for disadvantaged youth	Small initial positive effects on earnings that diminished to near zero over time[19]
Upward Bound initiative to help disadvantaged youth prepare for, enter and succeed in college	Weak or no effects on postsecondary education[20]

Several of the above studies identified small subgroups for whom the intervention showed promising effects, indicating areas for possible program refinement and further study. Importantly, rigorous evaluation can elucidate the true effects of programs and interventions, providing valuable information on what *does not* work to allow further learning about what *does*. Much can be learned from rigorous research to help develop more effective programs. In numerous areas of policy, a shift to using rigorous research to inform decision making in policy and programming can improve investment returns and result in interventions that produce significant, meaningful improvements for children and families.

Evidence-based programs are not always available to inform policymaking. Where evidence-based programs exist, priority funding for these programs will maximize positive outcomes. In areas with no existing evidence-based programs, policymakers can pilot innovative programs that are based on existing theory and research, followed by rigorous evaluation to learn whether they work and how they might be improved. (See the brief "Designing Better Pilot Programs: 10 Questions Policymakers Should Ask" in this report.)

What are the types of study designs?

While RCTs are the gold standard in research, they may also be time consuming, logistically challenging and expensive. As a result, less rigorous evaluation methods make up approximately 90 percent of evaluation studies. Such designs can be useful in generating *hypotheses* about what works, and indeed are a good first step in determining which interventions are ready to be tested more rigorously. They do not provide strong evidence of effectiveness, however, and unless they are used carefully they may easily lead to erroneous conclusions.

Summary of Selected Study Designs, from Most to Least Rigorous	
Study Design	**Essential Components**
Randomized controlled trial	• Comparison of two or more interventions or one intervention and a control group • Random assignment of recipients to interventions to ensure that groups are equivalent
Quasi-experimental design with observably equivalent groups	• Comparison of two or more interventions or one intervention and a control group • Groups are highly similar in all key characteristics • Data are preferably collected before and after intervention
Comparison-group study with non-equivalent groups	• Comparison of two or more interventions or one intervention and a control group • Groups are not equivalent in key characteristics, though statistical procedures may be used to "control for" group differences
Pre-post study	• Comparison of individuals' pre-intervention and post-intervention scores on relevant measures to identify change over time • Does not account for change that would have happened anyway, regardless of intervention participation
Outcome metrics	• Review of participant outcomes without reference to a control or comparison group • Does not provide a baseline from which to measure success

What are less rigorous study designs?

Commonly used but less rigorous study designs include comparison-group studies, pre-post studies and outcome metrics, each of which is described briefly below with an example.

Comparison-group studies include two or more groups that are <u>not</u> equivalent in key characteristics.

In these studies, statistical procedures (such as propensity scores or covariate analyses—see glossary for definitions of these terms) can be used in an attempt to "control for" group differences. Findings cannot always be trusted with a high level of certainty, however, as unobserved group differences may exist (e.g., motivation to change, as in the example below). Consider the following results of two study designs examining a career academy intervention that attempts to improve high school graduation rates (see figure 1).[21]

1. *Nonrandomized comparison group*: A comparison group was selected from a nationwide population of like students from

similar schools, with statistical procedures used to control for observable group differences. Results indicate that the career academy intervention has a large effect on high school graduation rates.

2. *Randomized Controlled Trials*: Students who volunteered for the career academy were randomly assigned to either the intervention or control group. With this research design, the intervention effect disappears—the two groups had comparable graduation rates.

Problem: In the nonrandomized design, the intervention and comparison groups were not equivalent. Students who volunteered for participation in a career academy were those who already had motivation to graduate and succeed in school, while those from the nationwide sample include a mix of motivation levels. Without the RCT, policymakers might conclude erroneously that this program was effective at increasing graduation rates and consequently spend valuable intervention dollars on a program that does not work.

Figure 1: Impact of Career Academies on High School Graduation Rates

Source: Data provided by James Kemple, MDRC Inc.
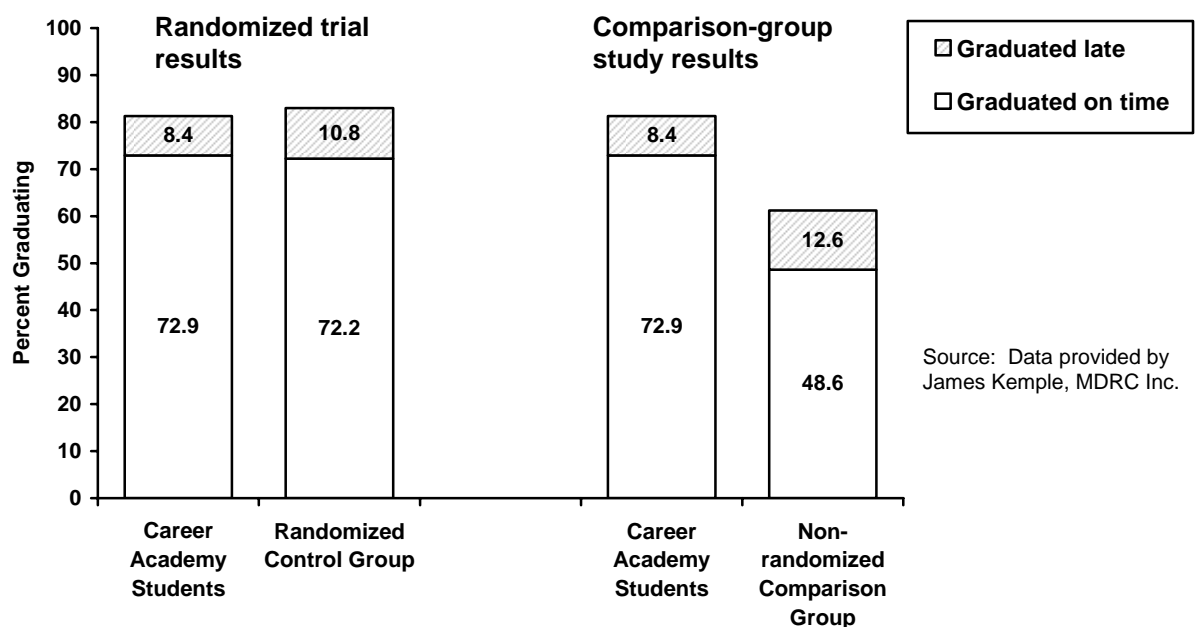
Pre-post studies use intervention recipients as their own control group by comparing pre-intervention scores on relevant measures with the scores received after the intervention is complete.

This design ensures group equivalence on key characteristics but fails to account for the passage of time or for other interventions and events that may have taken place concurrently. Consider figure 2 from a study on a national job training program.[22] Looking at the pre-post scores of the intervention group alone, it appears as if this program increased the earnings of young males. With no control group to serve as a comparison, one might mistakenly conclude that the program was successful. In fact, as compared with the control group, program participants actually had a smaller increase in earnings.

Outcome metrics may be used without reference to a control or comparison group.

This design provides outcome data but fails to provide any baseline from which to measure success. Consider the adult outcomes for individuals who participated in the Perry Preschool Project:[23,24]

- 35 percent did not finish high school or complete a GED.
- 32 percent had been detained or arrested.
- 57 percent of females had out-of-wedlock births.
- 59 percent received government assistance (e.g., welfare).

Was this program effective? By themselves, these numbers suggest that a large number of program participants had troubling adult outcomes: there is no frame of reference for comparison. Outcomes *compared to a control group* show large positive effects, however (see figure 3).

Figure 2: Job Training Partnership Act: Impact on Earnings of Male Youth

Figure 2B

Mean Earnings, by Quarter: Male Youth Non-arrestees
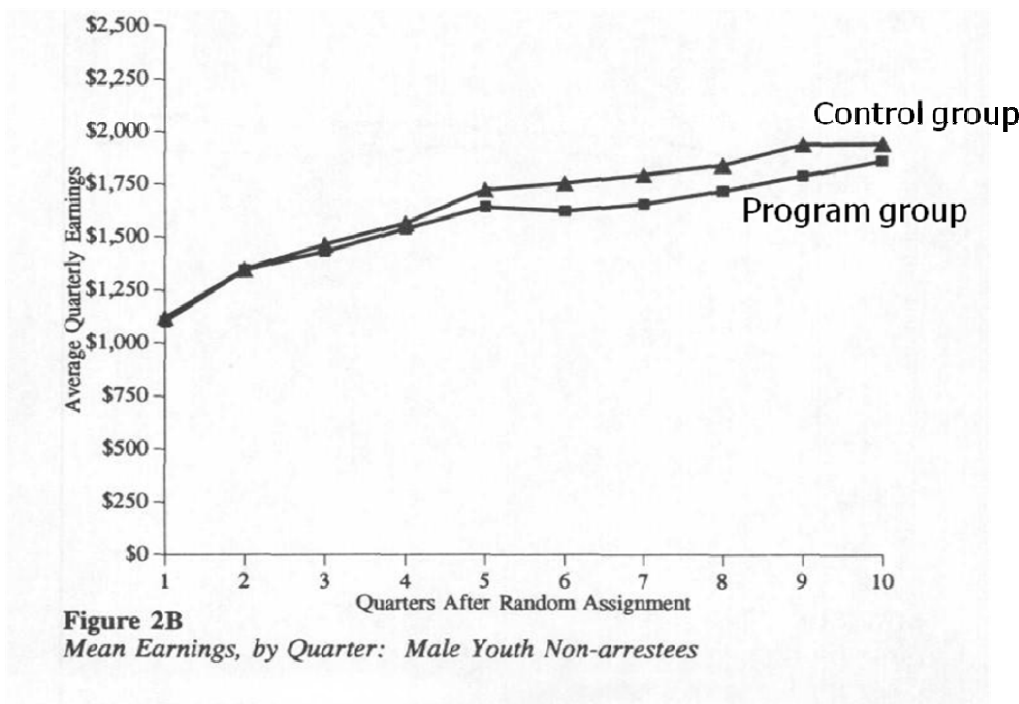
Source: Bloom et al., 1997

Figure 3: Impact of Perry Preschool Project on Life Outcomes
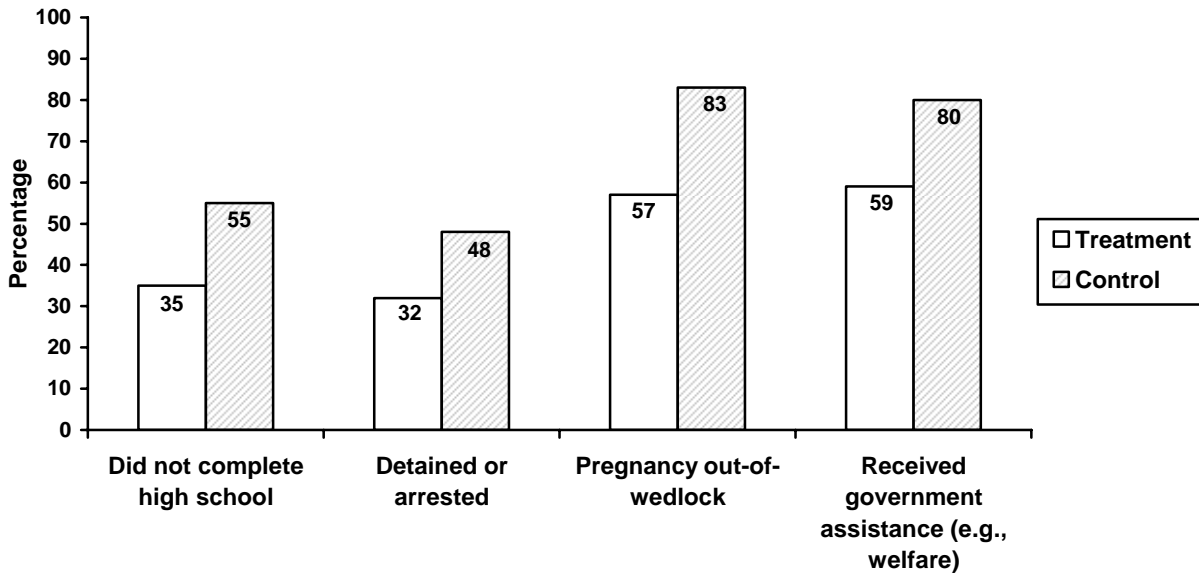
Source: Schweinhart, L. J., Barnes, H. V., & Weikart, D. P., 1993

Again, outcome-metric study designs can be valuable, both in providing preliminary hypotheses about program effectiveness and in answering other types of research questions (e.g., questions about risk factors or development of a problem over time). However, obtaining conclusions about program impact requires stronger study designs.

How do the strongest study designs differ from others?

Well-designed randomized controlled trials provide the strongest, most reliable results about program effects and are therefore the best design for informing policy decisions. RCTs are characterized by:

1. Comparison of two (or more) fundamentally different interventions (or one intervention versus "services as usual"), and
2. Random assignment of recipients (individuals, groups, towns, etc.) to the different interventions in order to balance both the observed and unobserved differences among the groups (i.e., to ensure that the groups are equivalent).

Not all RCTs are created equal. Even with random assignment to groups, there are design flaws that can bias study findings. When reading a study, consideration of the key design elements presented on page 12 will help determine how much confidence to place in the results or how likely it is that the study produced valid evidence of program effectiveness.

The best alternative to RCTs is a quasi-experimental design with observably equivalent intervention and comparison groups. This is a study design in which:

1. The intervention is compared with one or more observably equivalent control or comparison conditions;
2. Subjects are not randomly assigned to study conditions; and
3. Data are preferably collected at pretest and posttest.

The groups in this type of a study should be highly similar in key characteristics, including their predicted degree of motivation. Preferably, the comparison groups should be selected prospectively (i.e., before the intervention is administered).

Key Elements of a Well-Designed Randomized Controlled Trial

1. The study clearly describes the intervention (e.g., who did what to whom, and for how long).
2. If appropriate, the study randomly assigns groups (e.g., classrooms, counties), not just individuals within those groups (e.g., students, county residents).
3. The study has an adequate sample size—one large enough to detect meaningful effects of the intervention. One can feel confident that the sample size is adequate if statistically significant effects were identified or if a power analysis was conducted (refer to the glossary for definitions of *statistical significance* and *power analysis*).
4. The study shows that the intervention and control groups were similar in key characteristics prior to the intervention.
5. Few or no control group members participated in the intervention or otherwise benefited from it (i.e., no "crossover" or "contamination").
6. The study obtained outcome data for a high proportion of the sample members originally randomized (i.e., there is low attrition).
7. The study reports outcome data even for those in the intervention who do not complete (or even start) the intervention (i.e., "intention-to-treat approach").
8. The study uses outcome measures that are valid (i.e., highly correlated with the true outcomes the intervention is designed to affect). Examples include educational or psychological tests whose validity is well established or objective measures of the outcome (e.g., arrest rates). Self-reported outcomes are preferably corroborated by independent and/or objective measures. Outcome measures should not favor the intervention over the control group, or vice versa.
9. Where appropriate, evaluators are kept unaware of who is in the intervention versus the control group (i.e., evaluators are "blinded").
10. The study measures key policy or practical outcomes that the intervention seeks to affect (e.g., reduction in instances of partner violence), not just surrogate outcomes (e.g., changes in attitudes about violence).
11. The study preferably obtains data on long-term outcomes of the intervention (e.g., a year or longer after the intervention ends).
12. If the study claims that the intervention is effective, it should report:
    a. the size of the effect (i.e., "effect size") and
    b. statistical tests showing that the effect is unlikely to be due to chance. If groups (e.g., classrooms) instead of individuals were randomized, hierarchical tests should be used (refer to the glossary for definition of *hierarchical tests*).
13. The study reports the intervention's effect on *all* of the outcomes that the study measured.
14. Preferably, the study evaluated the intervention in the real-world community settings and conditions where it would normally be implemented.
15. A study's claim that the intervention's effect on a subgroup (e.g., Hispanics) differs from the effect on the overall population should be treated with caution until corroborated in one or more additional studies.

What policy actions support rigorous research design?

The principles of evidence-based policy suggest that the following strategies can strengthen outcomes and maximize investment returns on publically funded initiatives:

1. *Support programs that work.* The above information on study designs should equip policymakers and other stakeholders to begin evaluating evidence on program effectiveness. There are also many organizations that have critically evaluated study findings and ranked programs based on their level of proven effectiveness. Some are listed in Appendix III, the resource section of this briefing report. Once effective programs have been identified, it is up to the policymakers, agency officials and program administrators to support their implementation.

Strategies for supporting evidence-based programs include the following:

- Funding widespread implementation only for programs with proven effectiveness;
- Providing strong incentives and assistance for service providers to adopt research-proven interventions;
- Funding infrastructure to ensure programs are delivered effectively and with fidelity to the program model;
- Monitoring program implementation and outcomes on an ongoing basis to support continuous quality improvement and ensure that programs are meeting desired goals; and
- Ensuring that promising new ideas are piloted and tested.

2. *Build the evidence for new and/or untested programs using pilot programs.* There are plenty of "good" ideas that appear likely to be effective and find their way into programs. Careful piloting and testing of these programs before broad dissemination will provide opportunities for program enhancements and minimize dollars spent on ineffective services. To create such opportunities, policymakers

could consider allocating a small portion of funds toward the rigorous study of programs that show promise based on initial piloting and sound logic models (that is, the reasoning behind why a program is expected to work) but for which more evidence is needed before extensive replication. This will build the knowledge base about "what works" and increase the number of available evidence-based programs.

The following strategies for new program development can maximize the effectiveness of evaluation spending:

- Use RCTs whenever possible to evaluate the effectiveness (or "impact") of an intervention. If not, consider a well-matched comparison-group study (bearing in mind that careful consideration of group equivalence is key);
- Focus rigorous evaluations on only the most promising interventions. Though well-designed RCTs can sometimes be done at modest cost by using natural control groups such as waiting lists, they are generally more expensive to complete successfully than are less rigorous evaluation designs;
- Make sure that an intervention is well developed and well implemented before rigorously evaluating its effectiveness;
- Clearly outline, in advance, the tools and standards for measuring program success; and
- Be patient in awaiting results before making funding decisions about program replication and continuance: seeing the true program outcomes takes time. Participants must be recruited and receive the intervention and then have sufficient time for follow-up after the intervention is over (typically at least a year) to determine whether program effects are maintained over time. If funding decisions cannot wait for this process to be complete, erroneous decisions are likely.

3. *Use grant and/or contract mechanisms to encourage rigorous evaluations. Following are several possibilities*:

13

- Grants that include competitive priority for projects that include a rigorous (preferably randomized) evaluation;
- Grants that include absolute priority (i.e., requirement) for projects to include such an evaluation;
- Programs that sponsor an evaluation and require grantees to participate in the evaluation if asked;
- Programs that fund sheltered competitions to evaluate a specific model at several program sites with strong programs and capacity for rigorous evaluation; and
- Agencies that "waive" laws/regulations to allow demonstration projects and require rigorous evaluation.

Regardless of the policy area or challenge to be tackled, using strong evidence to inform intervention selection and implementation will enhance the likelihood of positive outcomes. Given limited funding resources, strategic support of proven programs is all the more critical to maximizing benefits. Where proven strategies do not exist, identification of promising interventions (based on pilot outcomes and solid logic models that show why the program is expected to be successful) will provide a starting point for limited initial implementation. Rigorous evaluation and iterative program improvements will yield new evidence-based practices, ultimately building a comprehensive menu of proven programs to enhance the well-being of North Carolina citizens.

[1] Olds, D. L., Eckenrode, J., Henderson, C. R., Kitzman, H., Powers, J., Cole, R., et al. (1997). Long-term effects of home visitation on maternal life course and child abuse and neglect: 15-year follow-up of a randomized trial. *Journal of the American Medical Association, 278*(8), 637-643.

[2] Olds, D. L., Henderson, C. R., Cole, R., Eckenrode, J., Kitzman, H., Luckey, D., et al. (1998). Long-term effects of nurse home visitation on children's criminal and antisocial behavior: 15-year follow-up of a randomized controlled trial. *Journal of the American Medical Association*, *280*(14), 1238-1244.

[3] Freedman, S., Friedlander, D., Lin, W., & Schweder, A. (1996). *The GAIN evaluation*: *Five-year impacts on employment, earnings, and AFDC receipt*. New York, NY: MDRC, Inc.

[4] Aos, S., Lieb, R., Mayfield, J., Miller, M., & Pennucci, A. (2004). *Benefits and costs of prevention and early intervention programs for youth*. Olympia, WA: Washington State Institute for Public Policy.

[5] Kolata, G. (2008, June 7). Tight rein on blood sugar has no heart benefits. *New York Times*.

[6] Manson, J. E., Hsia, J., Johnson, K. C., Rossouw, J. E., Assaf, A. R., Lasser, N. L., et al. (2003). Estrogen plus progestin and the risk of coronary heart disease. *New England Journal of Medicine, 349*(6), 519-522.

[7] Wassertheil-Smoller, S., Hendrix, S., Limacher, M., Heiss, G., Kooperberg, C., Baird, A., et al. (2003). Effect of estrogen plus progestin on stroke in postmenopausal women: The Women's Health Initiative randomized controlled trial. *Journal of the American Medical Association, 289*(20), 2673-2684.

[8] Beresford, S. A., Johnson, K. C., Ritenbaugh, C., Lasser, N. L., Snetselaar, L. G., Black, H. R., et al. (2006). Low-fat dietary pattern and risk of colorectal cancer: The Women's Health Initiative randomized controlled dietary modification trial. *Journal of the American Medical Association, 295*(6), 643-654.

[9] Grady, D. (2006, November 16). When blind faith in a medical fix is broken. *New York Times*.

[10] Kaplan, K. (2008, December 21). Vitamin supplements don't fight cancer, studies show. *Los Angeles Times*.

[11] Brown, D. (2005, April 19). Establishing proof: Some fifty years ago a baby-blinding epidemic confounded experts—until a pioneering study conclusively tied cause and effect, and enshrined clinical trials in medical practice. *Washington Post*.

[12] Brown, D. (2008, March 21). Vaccine failure is setback in AIDS fight: Test subjects may have been put at extra risk of contracting HIV. *Washington Post*.

[13] Kolata, G., & Eichenwald, K. (1999, October 3). Health business thrives on unproven treatment, leaving science behind. *New York Times*.

[14] Burtless, G. (1985). Are targeted wage subsidies harmful? Evidence from a wage voucher experiment. *Industrial and Labor Relations Review, 39*(1), 105-114.

[15] Petrosino, A., Turpin-Petrosino, C., & Finckenauer, J. O. (2000). Well-meaning programs can have harmful effects! Lessons from experiments of programs such as Scared Straight. *Crime and Delinquency, 46*(3), 354-379.

[16] Perry, C. L., Komro, K. A., Veblen-Mortenson, S., Bosna, L. M., Farbakhsh, K., Munson, K. A., et al. (2003). A randomized controlled trial of the middle and junior high school D.A.R.E. and D.A.R.E. Plus programs. *Archives of Pediatric and Adolescent Medicine*, *157*, 178-184.

[17] US Department of Education, Planning and Evaluation Service, Elementary and Secondary Education Division. (2003). *Third national Even Start evaluation*: *Program impacts and implications for improvement*. Cambridge, MA & Bethesda, MD: Abt Associates, Inc.

[18] Mayer, D. P., Peterson, P. E., Myers, D. E., Tuttle, C. C., & Howell, W. G. (2002). *School choice in New York City after three years*: *An evaluation of the school choice scholarships program.* Washington, DC: Mathematica Policy Research, Inc.

[19] US Department of Labor Employment and Training Administration, Office of Policy and Research. (2003). *National Job Corps study*: *Findings using administrative earnings records data.* Princeton, NJ: Mathematica Policy Research, Inc.

[20] US Department of Education Policy and Program Studies Service. (2004). *The impacts of regular Upward Bound*: *Results from the third follow-up data collection.* Washington, DC: Mathematica Policy Research, Inc.

[21] Kemple, J. J., & Willner, C. J. (2008). *Career Academies*: *Long-term impacts on labor market outcomes, educational attainment, and transitions to adulthood*. New York, NY: MDRC, Inc.

[22] Bloom, H. S., Orr, L. L., Bell, S. H., Cave, G., Doolittle, F., Lin, W., & Bos, J. M. (1997). The benefits and costs of JTPA Title II-A programs: Key findings from the National Job Training Partnership Act study. *The Journal of Human Resources, 32*(3), 549-576.

[23] Schweinhart, L. J., Barnes, H. V., & Weikart, D. P. (1993). *Significant benefits*: *The High/Scope Perry Preschool study through age 27*. Ypsilanti, MI: High/Scope Press.

[24] Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2004). *The High/Scope Perry Preschool study through age 40*. Ypsilanti, MI: High/Scope Press.

Brief 2
Designing Better Pilot Programs:
10 Questions Policymakers Should Ask

Kristopher Nordstrom

Executive Summary

The majority of pilot programs in North Carolina have failed to produce clear evidence of success or failure, making it difficult for legislators to determine whether to expand or discontinue the programs. Although the General Assembly has expressed a strong desire to obtain clear, objective evaluations of new programs, the design of pilot programs often makes quality evaluation impossible.

The goal of this memo is to help policymakers avoid the pitfalls that have undermined past pilot programs.

To ensure that new pilot programs will provide clear, useful results, policymakers should answer the following 10 questions when reviewing program proposals:

1. What is the problem that needs to be solved?
2. How does the program address the identified problem?
3. What is the cost of taking the program to scale if it is successful?
4. Is there a budget or spending plan?
5. What outcome criteria will be used to determine the program's success or failure?
6. What alternative programs or solutions might also address the problem?
7. Does the design of the evaluation allow for meaningful results?
8. Are there problems in the evaluation design that will affect validity?
9. Is there sufficient time to observe effects?
10. Is the sample size large enough to identify statistically significant effects?

With clearer results, policymakers will better be able to determine which programs work and which programs do not.

Introduction

Pilot programs are new initiatives implemented on a limited basis as a test or trial. As part of any pilot, program implementers should collect sound data to show whether or not the new program has potential to succeed on a larger scale or whether it should be discontinued.

The North Carolina General Assembly has demonstrated an admirable willingness to try out new initiatives by funding pilot programs. North Carolina's pilot programs have generally included provisions and funding for program evaluation. Unfortunately, these evaluations have often provided ambiguous results, making decisions on program expansion difficult for policymakers. The primary reason is that the programs and/or their evaluations have been designed in ways that inadvertently preclude meaningful assessment.

Common problems with pilot programs include the following:

- Unclear goals: what does it mean for a program to "work?"

- Unclear criteria: what measurements will be used to determine if a program is successful?

- No control group: results of the program are not compared against an independent group unaffected by the pilot program.

- Selection bias problems: sites that are in the program are systematically different from those that are not.

- An inadequate timeframe in which to observe outcomes: some pilot programs have been discontinued before results can be observed.

- An inadequate number of pilot sites: the number of sites is insufficient to produce meaningful data.

The goal of this brief is to help policymakers avoid the pitfalls that have undermined past pilot programs. Though this brief was written with pilot programs in mind, the recommendations could apply to the development of any new program.

When reviewing program proposals, policymakers should ask the following 10 questions. Answers should provide policymakers with the necessary information to assess program fit, feasibility, goals and evaluation design. Policymakers can then make informed funding decisions to ensure that new pilot programs are well designed and will provide unambiguous results. With clear results, policymakers will be able to determine which programs work and which programs do not, ensuring that taxpayer funds are directed to the best possible investments.

Question 1: What is the problem that needs to be solved?

Developing a clear problem statement is the first and most crucial step in the development of new pilot programs. Both program developers and policymakers should be able to articulate the nature, magnitude and distribution of the social problem targeted by a potential new program.

Development of the problem statement provides program developers a sense of direction and guides both implementation and evaluation. Furthermore, with a thorough understanding of the problems that the program seeks to alleviate, policymakers will be better able to weigh funding choices against competing claims on state resources.

Program developers and policymakers should avoid problem statements that define the solution to the problem or contain causal claims.[1] For example, consider the statement "children are dropping out because of a lack of laptops in the classroom." This statement makes a causal claim (that dropping out is the result of too few laptops) that might not be true and defines the solution (provide more laptops). A more useful problem statement would be "too many children are dropping out."

Question 2: How does this program address the identified problem?

Advocates for a new program or initiative should be able to explain clearly the theory or conceptual framework that suggests the program will solve the identified problem.[2] There should be a clear, logical and unambiguous relationship between the problem and the remedies that are to be applied to the problem.

Once the program's theory and conceptual framework are provided, the policymaker must critically assess the claims by asking the following:

- Do the program claims seem reasonable? If something sounds too good to be true, it probably is. The vast majority of successful programs make improvements at the margin.

- Is there existing research backing the program's claims? Legislators should consult with legislative staff to see if research exists.[3]

- Are there any scenarios that could cause this proposal to fail? Ask whether the parties responsible for implementing the program have considered possible pitfalls or roadblocks to implementation. What safeguards and contingency plans are in place?

Question 3: What is the cost of taking the program to scale if it is successful?

Pilot programs focus initially on a subset of the target population and a limited number of sites, in part to keep total costs manageable. It might be relatively easy for the state to find money to fund a pilot program. What will happen, however, if the program is successful? Will the program still be affordable if it is offered to the entire target population?

For example, consider a pilot program that delivers a new service to four schools at a cost of $1 million. There are approximately 2,400 schools across North Carolina, so expanding to a full-scale program statewide would cost $600 million. However, costs would be substantially lower if the target population included only high poverty schools or those schools tailored to students with special needs.

North Carolina legislators should consult with the Fiscal Research Division to determine how much a full-scale program would cost. If a full-scale program is something that would be cost-prohibitive, there is little point in conducting the pilot program unless the ultimate goal is limited replication.

Question 4: Is there a budget or spending plan?

Policymakers should examine budgets to assess whether or not the proposed pilot program:

- Has been thoroughly planned,
- Aligns spending to the program's stated goals, and
- Includes the resources necessary for successful implementation and evaluation.

A well-crafted, reasonable budget is an indication that thought has been given to how the new program will be executed. A vague, poorly crafted budget (or worse, no budget!) may indicate that the program has undergone only minimal planning. A detailed budget allows policymakers to assess whether the spending plan aligns with the program's stated goals (i.e., program priorities are well funded) and includes the resources necessary for successful implementation and evaluation. Two potentially critical planning and budget items are commonly neglected in pilot program budgets:

1. Professional development for program staff, and
2. Program evaluation.

Consider a pilot program focused on implementing a new drug-treatment method. The staff implementing the program might require training and careful supervision to introduce the new procedures into practice. The expense might be significant, but it could be crucial to the successful implementation of the program.

Similarly, a program evaluation that provides reliable results can be expensive. However, without proper evaluation, the pilot program will likely generate ambiguous data.

Bear in mind that a quality budget plan is an indication of how well the program is likely to be implemented, but it is not necessarily an indication of how effective the program will be in achieving its objective. Even exceptionally well-run programs might not have a discernable impact on a problem.

Question 5: What outcome criteria will be used to determine the program's success or failure?

Policymakers should establish in advance the criteria for determining the success of a pilot program. What will a successful program accomplish? How will results be measured? How large does the program's effect need to be? The criteria for evaluating a program should be

18

objective, measurable, unambiguous and relevant to the program's goals.

For example, clear criteria for an education pilot program could include improvement in student test scores, dropout/graduation rates and teacher turnover. In addition to looking at overall test results, legislators might consider equity measurements. A new program could show great increases in test scores overall, but effects could vary widely among different groups of students. This might be a perfectly acceptable result. However, if the sample size is large enough, policymakers are encouraged to examine results for various subcategories of students and, to the extent possible, define the levels of disparity that would be considered acceptable.

Selecting the outcome criteria will allow policymakers to identify the types of data needed to evaluate the program. At the end of the evaluation, what specifically do policymakers want to know? Ideally, the pilot program will show two things:

1. Program participants experience a specific outcome; and
2. Similar persons that are not exposed to the program do not experience that outcome.

Program implementers should work with the relevant state agencies to ensure that they can get the data needed to evaluate program outcomes. If an agency lacks the capacity to gather the required data, policymakers must decide whether to provide the agency with the necessary capacity to gather it or to identify alternative criteria that will still show meaningful results.

Question 6: What alternative programs or solutions might also address the problem?

For any identified problem, there are likely programs, products or services being tried in other states to address the problem. It is important that policymakers consider those and any other relevant alternatives before choosing to appropriate state funds for a pilot program. There might be alternatives that provide a greater likelihood of success or can achieve similar ends at a lower cost.

Program advocates should disclose what alternatives exist, when asked. Legislative staff can research potential alternatives on behalf of interested legislators.

Question 7: Does the design of the evaluation allow for meaningful results?

The most common reason pilot programs fail is that their evaluation designs do not allow evaluators to demonstrate the program's results clearly. Most rigorous evidence falls into one of two design categories: a randomized controlled trial or a comparison-group study with equivalent groups.

Randomized controlled trial: In this design, participants (e.g., individuals, schools, communities) are randomly assigned to either a treatment group (participation in the new pilot program) or a control group (no participation in the pilot program). Evaluators use random assignment to form two equivalent groups in the most objective way possible.[4] This structure is the most rigorous technique to determine whether the observed outcomes are a product of the program, rather than a product of other factors. Few pilot programs in North Carolina have included randomized controlled trials. However, most could have been designed as such with additional planning. Randomized controlled trials should be used whenever feasible.

Comparison-group study: In a comparison-group study, there are still control and treatment groups, but participants are not randomly assigned to the groups. Instead, participation in the two groups is based on observable characteristics (e.g., demographics) and evaluators strive to make the groups as similar as possible. With a comparison-group study, it is more difficult to demonstrate with confidence that an observed effect is caused by the pilot program. However, such studies can provide tentative support for a program and might be the only option when implementation of a randomized controlled trial is not feasible.

There are many other study designs that do not allow for meaningful evaluation of a pilot program. These designs provide indications of potential program effects rather than conclusive findings.

They might help policymakers decide whether a more conclusive evaluation would be worthwhile.

"Pre-post" studies: Participants are assessed before and after the intervention. Changes are assumed to be caused by the intervention.

Poorly designed comparison-group studies: A comparison group is selected but is not closely matched with the treatment group on all relevant variables.

Anecdotal evidence / satisfaction: Selected testimony or a measurement of participant satisfaction is presented as evidence that a program is working. Anecdotes do not rise to the level of evidence. When only anecdotal evidence is presented, it is likely a sign that the program lacks meaningful evidence of success.[5]

Question 8: Are there problems in the evaluation design that will affect validity?

Ideally, a new pilot program will produce results that have high validity. That is, the program will adequately demonstrate that:

- The intervention is actually causing the desired outcome (internal validity), and
- The program is replicable, producing similar results in different settings (external validity).

Randomized controlled trials inherently minimize most threats to validity. However, since few pilot programs in North Carolina are evaluated with randomized controlled trials, policymakers should examine evaluation results for some of the following common threats to validity.[6]

Self-selection bias: A common design flaw of North Carolina's pilot programs and of pilot programs generally is self-selection bias. That is, pilot programs are conducted only in places that have expressed a desire to participate in the program. The problem is that participants' decisions to participate may be correlated with traits that affect the study results. For example, schools that choose to participate in a pilot program might have teachers with higher levels of motivation than schools that choose not to participate. As a result, it may appear that the pilot program is working when the results are really just

a reflection of the differences in teacher motivation.

Non-representative samples: Pilot programs are tested with small samples of participants, and if deemed successful, they are scaled up to include a larger population. Often, however, the participants in the pilot program are not representative of the broader population that would be served under the full-scale program. For example, many pilot programs in North Carolina are introduced in the smallest counties or the most economically disadvantaged areas. As a result, it is difficult to generalize the results. Will the program work across other counties in the state? Certain programs might be more effective in rural than urban areas, or the program might have differing effects on different minority groups. Policymakers can be more confident that the pilot program is replicable if the sample participants are as representative as possible of the total population to be served.

Question 9: Is there sufficient time to observe effects?

Meaningful evaluation may require substantial time to observe a program's effects. Educational programs that involve new ways of teaching, for example, might require a one- or two-year ramp-up as teachers adapt to the new teaching method.

Other programs might be focused on long-term effects. In the case of a substance abuse program for young children, for example, the evaluation must take time to wait for long-term observations of substance use.

Additionally, time is required to gather enough observations to determine if initial effects are replicated and maintained. If the observed effects are replicated year-after-year, it is more likely that they are a result of the program intervention. If individual outcomes last over time, they can be considered meaningful change.

Policymakers should ensure that a pilot program has sufficient support to allow time for meaningful evaluation. If the plug is likely to be pulled before the program is able to produce results, then it is not worth pursuing.

Question 10: Is the sample size large enough to identify statistically significant effects?

In order for study effects to be statistically significant, the study must have a sufficiently large sample size. The required sample size varies based on what unit of study is chosen (e.g., students, classrooms, schools, districts). The table below presents rules of thumb on sample sizes for educational pilot programs.[5]

Unit of Study	Sample Size (includes both control and intervention groups)
Students	300
Classrooms	50 - 60
Schools	40 - 50
Districts	15 - 20

Actual numbers required will vary from study to study. Depending on the program and outcomes assessed, more or fewer units of study might be required. Fiscal Research analysts can work with parties designing new pilot programs to ensure that the program will include a sufficient sample size to provide meaningful results.

Conclusion

Policymakers can use these 10 questions to guide their investments toward better pilot programs, which in turn will support the development of programs with better outcomes and better use of taxpayer dollars. However, simply asking the questions is not enough.

Policymakers should insist upon pilot programs that are designed as randomized controlled trials whenever possible. A randomized controlled trial means that certain groups (such as counties or school districts) will be receiving the pilot program intervention (the treatment group) while others will not (the control group).

Policymakers should refrain from insisting that their districts be included in treatment groups. This is important to ensure that results are not skewed. Also, being in the control group can be beneficial. Not all pilot programs are helpful. More importantly, control groups are necessary to develop new programs that will eventually benefit all members of a target population. A pilot program that generates actionable data is far more important than having a poorly designed program placed in a home district.

Additionally, policymakers should allow time for pilot programs to reach their full implementation and demonstrate program effects. Acting too early might result in the abandonment of programs that are actually working.

A combination of smart policy design and a measure of political restraint is required for the development of quality pilot programs. With better pilot programs, policymakers can make smarter investments in new programs and place North Carolina at the forefront of policy innovation.

_____

[1] Bardach, Eugene. (2000, September 1). *A Practical Guide for Policy Analysis*: *The Eightfold Path to More Effective Problem Solving*. CQ Press, 5.

[2] Program advocates are encouraged to develop a theory of change and logic model for their projects. Additional information on these topics can be found at the Centers for Disease Control Web site (www.cdc.gov/eval/resources.htm#logic%20model) and the W.K. Kellogg Foundation Web site (www.wkkf.org/default.aspx?tabid=75&CID=281&NID=61&LanguageID=0).

[3] Members desiring to do their own research may wish to begin their search at the What Works Clearinghouse (ies.ed.gov/ncee/wwc/overview/) or the Promising Practices Network (www.promisingpractices.net/).

[4] Myers, D. & Dynarski, M. (2003). *Random Assignment in Program Evaluation and Intervention Research*: *Questions and Answers*. Washington, DC: Institute of Education Sciences.

[5] US Department of Education Institute of Education Sciences National Center for Education Evaluation and Regional Assistance. (2003). *Identifying and Implementing Educational Practices Supported by Rigorous Evidence*: *A User Friendly Guide*. Washington, DC.

[6] Information for this section is based on lecture notes from Christina Gibson-Davis's Qualitative Evaluation Methods course (Pubpol 313) at Duke University, taken Spring 2005.

Using Research to Drive Public Policy: The Case of the North Carolina Graduated Driver Licensing System

Robert Foss

In 1997 North Carolina was one of the first two states in the nation to adopt a Graduated Driver Licensing system (GDL) to reduce motor vehicle crashes among young novice drivers. This policy was driven heavily by scientific research. At that time, motor vehicle crashes were the leading cause of death among teenagers. The adoption of GDL led to a sharp decline in teenage driver crashes and a decline in the motor vehicle death rate among teenagers.

The adoption of GDL is an example of how the North Carolina General Assembly used the best available scientific evidence to enact policies that have proved helpful in addressing a serious threat to teenagers and those with whom they share the roads. Careful research conducted in North Carolina since 1997 has helped establish the GDL system as an evidence-based program. As a result, 46 other states have adopted GDL. The system is unique in the history of traffic safety in its ability to reduce crashes among the youngest drivers. Studies of the best GDL systems report crash rate declines of nearly 40 percent among 16-year-olds and 20 percent among 17-year-olds.[1] This brief describes how GDL was adopted in North Carolina and offers it as a model of how research can inform future policymaking.

How the Program Started

Crashes, injuries and deaths among teenage drivers were and continue to be serious problems. What we have learned from the GDL program can inform continued efforts to reduce crashes, injuries and deaths on North Carolina's roads. Motor vehicle crashes account for about 35 to 40 percent of deaths from all causes among teens 15 to 19 years of age in the United States. The teen driver problem is particularly acute among the youngest drivers. Among 16-year-old drivers, the crash rate per mile driven is about 1.5 times greater than for 17-year-olds, 2.7 times greater than for 18-year-olds, 3.9 times that of 19- and 20-year-old drivers, and nearly 10 times that of 30- to 59-year-old drivers.[2] Moreover, motor vehicle death rates per licensed driver among all drivers older than 17 years declined from 1976 to 1996, reflecting the effects of a variety of safety improvements. Yet the per driver death rate among 16-year-olds increased considerably, apparently indicating an inability of novice drivers to cope with an increasingly complex driving environment.[3]

The first steps toward a policy approach to dealing with a complex problem like motor vehicle crashes are to document the size of the problem and then, based on an understanding of its nature, to determine whether a policy might be expected to have a beneficial effect and be feasible to implement. The better a problem is understood, the more likely it is that we can determine where to intervene and what sorts of interventions might reasonably be expected to reduce the problem. The North Carolina Child Fatality Task Force, established by the North Carolina General Assembly and the North Carolina Governor's Highway Safety Program, worked to publicize the number of fatalities associated with teenage-driver automobile crashes. This helped put this issue on the public's agenda. The task force asked researchers at the University of North Carolina Highway Safety Research Center for guidance on how to reduce the teenage crash rate. Analysis of when, where and why teen drivers were involved in crashes was an invaluable first step toward finding potential solutions.

Analyses of crash data revealed patterns that suggested possible interventions. First, the crashes were much more highly concentrated among the youngest and least experienced drivers, something that—surprisingly—was not previously understood. Teens were lumped together and seen as uniformly poor drivers. Analysis of crashes by single year of age revealed quite a different picture. Second, crash types suggested that lack of driving "savvy" or wisdom, rather than deliberately risky or foolish behavior (e.g., drinking, thrill-seeking), was the most common factor associated with crashes among these least experienced drivers. More specifically, crashes among novice drivers appeared to result from two things: (1) limited ability to accurately judge risk and opt for appropriate actions and (2) the impulsive behavior style that is characteristic of adolescents. Third, the most dangerous conditions for young novices involved nighttime driving, which is often "recreational" and involves multiple teenage passengers. The largest numbers of young driver nighttime crashes occur between 9 p.m. and midnight, not during the post-midnight hours typically thought of as high risk.

Once specific crash patterns were identified, researchers at the Highway Safety Research Center looked for evidence-based approaches that could address the identified phenomena (inexperience, impulsiveness and nighttime crashes). Researchers identified two independent evaluations of a graduated licensing program in New Zealand.[4,5] This program reduced teen driver crashes and injuries by providing young beginning drivers with substantial practical driving experience under the safest possible conditions, moving drivers through various levels with limited driving privileges before full licensure. New Zealand's GDL was based on evidence that with a year or two of experience, most people learn to drive reasonably safely. GDL changed the driver licensing approach to make the initial year or two of driving—when most of the essential learning takes place—safer for inexperienced drivers by specifying the conditions for when and how they are allowed to drive, depending on their experience. In brief, GDL tries to ensure that teens have plenty of experience obtained under realistic but reasonably safe conditions prior to being allowed to drive by themselves.

A number of studies also showed that restricting nighttime driving for young drivers led to substantial reductions in crashes, injuries and fatalities during restricted hours.[6,7] Given that the large majority of nighttime crashes among young drivers occur between 9 p.m. and midnight, a driving restriction during these early hours was needed to effectively reduce crashes among this age group. Restrictions beginning after 11 p.m. or midnight can address only a small fraction of the risk to the novice driver population. Although the risk *per trip* is much higher after midnight, the greatest safety gains come from starting the restrictions earlier, when most night driving among this age group occurs.

Other approaches that looked promising to some, such as improving the content of driver education classes, were not supported by research findings and did not appear to hold the promise of moving to a GDL system.

Based on this evidence, several groups, including the Child Fatality Task Force and the Governor's Highway Safety Commission, recommended enactment of a GDL system for North Carolina. When legislation was first introduced, late in the 1994-1995 legislative session, there was some opposition, resulting from inadequate

understanding of what GDL involved and why. During the following 20 months a concerted effort was made to ensure that the public and legislators understood both the need for GDL and how its combination of elements were carefully designed to address what was known about the nature of young teen driver crashes. The GDL bill was passed and ratified in April 1997 and took effect on December 1, 1997.

Before the enactment of GDL in North Carolina, an individual with virtually no practical driving experience could obtain an unrestricted license. Persons 15 years or older who had passed a mandatory driver education class, a vision test, a sign recognition test and a written driving test could begin driving if supervised; persons 16 years or older could begin with no supervision and with no practice besides the few hours obtained during a driver education class.

The North Carolina GDL system changed the licensing process by creating two preliminary licensing levels preceding a full unrestricted license. Both of these levels involve constraints on driving to limit the risks faced by inexperienced drivers, with specific restrictions regarding nighttime driving. North Carolina's

Levels of Licensing in the North Carolina GDL System (winter 2009)

Level	Requirements	Driving restrictions
Level 1	• At least 15 years old • Complete driver education • Pass written, sign recognition and vision tests	Driving allowed only while supervised by a designated adult No mobile phone use while driving* All occupants must wear seatbelts
Level 2	• Complete 12 months at Level 1 • No traffic convictions in the final 6 months of Level 1 • Pass road test	Unsupervised driving allowed 5 a.m. to 9 p.m.—driving at any other time must be supervised No more than one passenger younger than age 21** No mobile phone use while driving* All occupants must wear seatbelts
Level 3	• No traffic convictions in the final 6 months of Level 2	No mobile phone use while driving* All occupants must wear seatbelts

* Added December, 2006.
** Added December, 2002.

GDL system required beginning drivers who are at least 15 years old and younger than 18 years to first hold a Level 1 license (learner permit) for a full year before graduating to the next level. Level 1 allows driving only while supervised by a designated adult—typically a parent or guardian. After completing the final six months of Level 1 with no traffic violations and passing a road test, a driver may move to Level 2 licensure. Level 2 allows unsupervised driving from 5 a.m. to 9 p.m. and supervised driving at any other time. After completing at least six continuous months at this level with no traffic violations, drivers can graduate to a full, unrestricted license. Under GDL new drivers also must be at least 15 years of age to begin the process, at least 16 to move to level 2, and at least 16½ to move to a full, unrestricted license. Progression through these stages is achievement-based rather than merely a function of age. Equally important, young beginning drivers are required to earn the privilege of an unrestricted license by demonstrating safe driving behaviors during the first 18 months of licensure.

Did the program work? Establishing Evidence to Support Continuation and Expansion of the Graduated Driver Licensing System

Although evidence from New Zealand showed strong support for a system of graduated licenses, North Carolina did not implement a system identical to New Zealand's. Nor is the North Carolina driving environment comparable to that of New Zealand. Some policymakers were concerned about the relevance of research and practice from other countries to policy development in North Carolina. Others questioned what parts of the system were most important, especially as other states began considering similar legislation. Given the need for sound North Carolina evidence to support the continuation of this program, researchers at the Highway Safety Research Center examined the effectiveness of the GDL system using the strongest possible quasi-experimental research design for this kind of policy.

Many factors besides a new policy may contribute to changes in the rate of automobile crashes. Some of these include the price of gas, economic conditions, changes in traffic enforcement policies,

improvements in roadway engineering and maintenance, safer motor vehicles, and so on. To obtain the best estimate of the impact of the GDL system, researchers needed a valid comparison, or control group, to increase confidence that changes in the rate of young driver crashes were due to the new licensing system and not (entirely) to other factors. Since it was not possible to randomly assign new drivers to different licensing systems, researchers used a different kind of control group, in this case crash rates among 25-to-54-year-old drivers in North Carolina. Although not a perfect match, most of the factors other than the new licensing system that would influence crashes among younger drivers would also affect adult drivers. The only group for which conditions changed was young drivers. Researchers looked at crash rates among both young drivers and 25-to-54-year-old drivers before and after the implementation of the new licensing system. A significant change in crashes among young drivers following enactment of GDL, with no parallel change among adult drivers, would be good evidence that the change was a result of implementing GDL.

Analysis

Highway Safety Research Center researchers obtained data on crashes for all drivers of passenger vehicles (i.e., passenger cars, station wagons, vans, light pickup trucks, and sport utility vehicles) from the North Carolina Crash Data File. This file contains information on all reportable motor vehicle crashes (those involving a fatality, personal injury, or property damage valued at $1000 or more).

The analysis focused first on all crashes, then on subcategories based on severity (fatal, serious injury, or minor or no injury), time (day versus night), type (single versus multiple vehicle), alcohol involvement, and driving environment (more versus less urban counties). Crash severity was classified based on police officer reports of injury to occupants. In North Carolina, injuries are coded as fatal, incapacitating injury, visible minor injury, possible injury (complaint of pain with no visible injury), or no injury (property damage only). To compute rates, mid-year population estimates for North Carolina were obtained from

the US Census Bureau Web site. Driver license information was extracted from the North Carolina Driver History File.

The researchers compared 16-year-old driver crash data for December 1, 1998, through November 30, 1999 (hereafter referred to as 1999) with those from 1996 and 1997. Crash rates based on age-specific populations were computed to adjust for population growth from 1996 to 1999. To control for general crash trends that might reflect economic factors, special traffic safety initiatives or varying levels of enforcement, changes in 16-year-old driver crash rates were compared with those for drivers 25 to 54 years of age.

Results

The results were impressive. Across the state, crash rates declined sharply for all levels of severity among 16-year-old drivers after the GDL program was implemented. Comparing 1996 with 1999, fatal crashes declined 57 percent, and crashes with no or minor injuries decreased 23 percent. The benefit of the nighttime driving restriction, which lasts only six months for most young drivers, was similarly impressive. Compared with 1996, 16-year-old-driver crashes between 9 p.m. and 5 a.m. in 1999 were 43 percent less likely, whereas daytime crashes declined by 20 percent. Single-vehicle crashes declined somewhat more than multiple-vehicle crashes, reflecting the fact that inexperienced drivers are particularly prone to single-vehicle crashes (which almost always result solely from the driver's own actions and not those of others on the road).

When compared to the adult drivers the results are even more impressive. Among this older age group, the crash rate actually increased by almost 6 percent from 1996 to 1999. With the exception of alcohol-related crashes, population-adjusted rates for the older age group increased in every crash subgroup as well. This clearly indicates that the decreases among 16-year-old drivers did not result from a general downward trend in crashes. Adjusting for the overall crash trend, the crash rate among 16-year-olds decreased an impressive 27 percent from 1996 to 1999. Similar findings were obtained by researchers in Michigan, another early

adopter of a GDL program. Subsequent research in several other states has found similar beneficial effects of enacting a GDL system.[8] A more recent study examining the long-term effect of the North Carolina GDL system has documented a 38 percent decrease in crashes among 16-year-olds and a 20 percent decline among 17-year-olds, both of which are nearly identical to findings for the Georgia GDL program.

> This is the great advantage of policy approaches based on a scientifically sound understanding of a problem: where policy approaches are sustainable, there are certain and enduring results.

There has been a substantial amount of follow-up research on GDL looking at additional crash data, hospital admissions and associated costs as well as self-report surveys of parents and teens. Analysis of medical records showed that hospitalizations of 16-year-old drivers declined by 36.5 percent and that hospital charges for their care dropped by 31.2 percent, or $650,000 per year.[9,10] This research strongly underscores the effectiveness of GDL as well as widespread endorsement of the approach by parents and teens alike. Research in other states finds similar results. However, states with weaker GDL systems (e.g., shorter learner periods, inadequate night and passenger restrictions) obtain smaller crash reductions.

Most impressively, the effect of GDL has persisted over time. Because it is a policy approach that instills a permanent change in how an important process operates—the role of the licensing system in developing savvy drivers—the effects do not require a continual devotion of extra resources for education or enforcement efforts to be sustained. This is the great advantage of policy approaches based on a scientifically sound understanding of a problem: where policy approaches are sustainable, there are certain and enduring results. We believe that the success of the GDL effort in North Carolina has helped with subsequent efforts to improve highway safety. The same approach, using data and evidence to drive public policy, helped persuade the legislature to enhance the GDL system by adding a restriction on the number of youth allowed to ride with inexperienced drivers

(in 2002) and restrictions on the use of cell phones by young drivers (in 2006). In each case, initiatives were based on the analysis of crash data and scientific understanding of driver behavior. In both instances the effects of the policy changes have been studied carefully to provide the evidence needed to spread these policies to other states. It is not often that we can clearly demonstrate how the use of scientific evidence to inform policy can save so many lives. Our goal is for this kind of evidence-based policy to continue to be at the center of traffic injury prevention efforts and, through publications like this one, spread to other fields.

[1] Rios, A., Wald, M., Nelson, S., Dark, K., Price, M. & Kellermann, A. (2006). Impact of Georgia's Teenage and Adult Driver Responsibility Act. *Annals of Emergency Medicine, 47*(4), 369.e1-369.e7.

[2] Williams, A. F. (2003). Teenage drivers: Patterns of risk. *Journal of Safety Research, 34*, 5-15.

[3] Insurance Institute for Highway Safety (1998). Youngest drivers at risk. *Status Report. 33*(6), 1-2.

[4] Frith, W. J., & Perkins, W. A. (1992). The New Zealand graduated driver licensing system: National Road Safety Seminar, Wellington, New Zealand. *Seminar Papers, 2*, 256-278.

[5] Langley, J. D., Wagenaar, A. C., & Begg, D. J. (1996). An evaluation of the New Zealand graduated driver licensing system. *Accident Analysis and Prevention., 28*, 139-146.

[6] Williams, A. F., Lund, A. K., & Preusser, D. F. (1985). Night driving curfews in New York and Louisiana: Results of a questionnaire survey. *Accident Analysis and Prevention, 17*(6), 461-466.

[7] Preusser, D. F., Williams, A. F., Lund, A. K., Zador, P. L. (1990). City curfew ordinances and teenage motor vehicle injury. *Accident Analysis and Prevention, 22*(4), 391-397.

[8] Shope, J. T. (2007). Graduated driver licensing: Review of evaluation results since 2002. *Journal of Safety Research, 38*(2), 165-175.

[9] Margolis, L. H., Masten, S. V., & Foss, R. D. (2007). The effects of graduated driver licensing on hospitalization rates and charges for 16- and 17-year-olds in North Carolina. *Traffic Injury Prevention, 8*(1), 35-38.

[10] Foss, R. D. (2007). Improving graduated driver licensing systems: A conceptual approach and its implications. *Journal of Safety Research, 38*, 185-192.

Brief 4

Successfully Implementing Evidence-based Programs for Children and Families in North Carolina

Katie Rosanbalm, Michelle Hughes, Charisse Johnson, Rhett Mabry and Lauren Akers

Selecting and funding evidence-based programs can help achieve better outcomes for children and families. However, even the best evidence-based program will not yield good outcomes if it is not implemented well. This brief outlines the key components that policymakers and agency staff members might consider in drafting policy and promoting programs to achieve optimal outcomes for North Carolina. In particular, this brief discusses program fidelity and enumerates the core drivers for successful implementation. It then describes a unique, collaborative effort of public and private funders in North Carolina to fund and implement select evidence-based programs.

Importance of High-Quality Implementation

Imagine that North Carolina had a shortage of flu vaccines. To ensure the best possible outcomes for the population, responsible medical practitioners do not water down the vaccine so that it can be distributed to more people: they know that watered-down vaccines are ineffective. Instead, they target the full-strength vaccine to groups that are most likely to benefit from being vaccinated or those most at risk.

As in the flu shot example, watered-down social programs do not result in the intended outcomes for recipients. The National Implementation Research Network (NIRN) calls the typical orientation to social program implementation the "spray and pray" approach. Under "spray and pray," program developers, researchers and advocates "spray" program providers (e.g., social workers, teachers, nurses, etc.) with information and training in a program or practice, and then pray that providers implement it well. Service providers are on their own to institute and maintain practice changes, often assuming that a small dose of a good program is almost as good as the full dose.[1]

Considerable research indicates that the opposite is true: merely disseminating information does not produce either changes in practitioner behavior or benefits to consumers. Furthermore, full doses are required to produce positive outcomes for recipients. Without ongoing support, monitoring and reinforcement to help practitioners solidify their skills and knowledge, they will not be successful in either implementing a program well or in maintaining changes in practice.[1]

As an example, NIRN highlights a meta-analysis by Joyce and Showers[2] that summarizes research on training public school teachers. The results of different types of training and follow-up are shown in table 1. Across studies, as training components grew to include technique demonstration, practice and feedback, teacher knowledge and skill

demonstration improved substantially. However, actual classroom use of the new technique was achieved only with all of these components plus on-the-job coaching.

It is not sufficient simply to select a program model and share it with providers. Even with a great deal of traditional training, this approach is unlikely to achieve change in actual practice. Change in an implementing organization's underlying infrastructure is often required to prevent practitioners from drifting back to old practices.

Table 1: Success Rates in Improving Teacher Practices

	Domain of Teacher Competence		
Training Components	**Knowledge**	**Skill Demonstration**	**Use in Classroom**
Theory and Discussion	10%	5%	0%
…+Demonstration	30%	20%	0%
…+Practice and Feedback	60%	60%	5%
…+Coaching in Classroom	95%	95%	95%

Joyce & Showers, 2002

Program Fidelity

Program fidelity refers to how well a program is implemented according to established standards. Research on implementation of evidence-based programs shows that fidelity to core program elements is critical to success.[1,3] Core program elements (which may differ across programs) often include the following:

- Mechanism of service delivery (e.g., home versus office visits, lecture versus group discussion);
- Frequency and intensity of service delivery;
- Chronology of the service components;
- Staff education and professional credentials;
- Staff training requirements; and
- Substantive elements of the programmatic intervention.

Changing any of these components can adversely affect program outcomes. Indeed, suboptimal implementation of an evidence-based program may not simply reduce intervention effectiveness, it may actually cause harm to recipients. For example, the Strengthening Families Program has been identified as a model program for reducing youth substance use and delinquency and is currently implemented statewide in several states. However, a recent randomized trial evaluating a Strengthening Families Program found that, in contrast with previous evaluations at other sites, the intervention had slightly *detrimental* effects on child reports of negative peer associations and on family supervision and bonding.[4] In other words, while the goal of the program was to decrease child affiliation with delinquent peers and improve family relationships, the outcome at this site showed change in the opposite direction. The study identified major challenges in program implementation, including poor fidelity and high

staff turnover, which were likely key reasons that the program failed to achieve the positive effects documented by other Strengthening Families Programs. As a result of these findings, the authors stressed the critical need for sufficient infrastructure prior to program dissemination.

Challenges to delivering a program with fidelity typically include limited financial and human resources, incompatible or unrealistic community expectations, and lack of practitioner confidence in or thorough understanding of program principles. Policymakers may want to ensure that communities and program providers have considered these issues prior to program funding and implementation. If community leaders cannot deliver an evidence-based program with fidelity, they likely would be best served by focusing first on capacity building or by selecting a different program model.

Core Drivers for Successful Implementation

As summarized by NIRN,[1] successful replication of evidence-based programs requires a range of supports (or "scaffolding") for the community-based agency implementing the program. Agencies and program providers may need assistance with some or all of these scaffolding components in the form of funding, oversight and statutory authority. Legislation that includes feedback and reporting components, such as annual reports to policymakers and other funders and stakeholders, will ensure that evidence-based programs are staying on track and will highlight opportunities for program improvement, expansion of services, or both.

Essential infrastructure and scaffolding include the following:

- *Assistance with community and agency planning*: To ensure the greatest impact possible, programs must address specific community needs and must be a good "fit" for the target population, the community at large and the agency that is implementing the program. Additionally, the implementing agency must have sufficient capacity to execute the program effectively. Capacity includes resources, staff expertise and commitment from

leadership. Agency leaders and program advocates from within the agency and beyond will benefit from building both internal agency capacity and stakeholder buy-in (through collaboration and problem-solving) prior to program implementation.

- *Staff selection*: To achieve optimal program outcomes, staff members of the implementing agency must carry out program components with expertise, confidence and a solid understanding of how the program works. They should be adept at interacting with program recipients, show good judgment and a willingness to learn new skills, understand the value of evaluation and be willing to advocate for the program. Some programs require staff members to have specific credentials or experience. Others are developed so that a broader pool of providers can deliver the program.

- *Pre-service and in-service trainings*: Agencies implementing an evidence-based program likely will require instruction by a certified, quality trainer of the program to ensure that staff members are comfortable and knowledgeable about aspects of the program, such as its background, theory, philosophy, values, strategies and rationales of key practices. In addition, staff members must have the opportunity to learn new skills and practice them in a safe environment in which they can receive constructive feedback. Training for program staff members is typically provided by the program's national office or by the original program developer, although this varies by program. Ideally, ongoing in-service trainings provide support and education for program staff throughout the life of the program.

- *Ongoing consultation and coaching*: Coaching in the context of evidence-based programs is broader than traditional supervision. It is a supportive, strengths-based strategy to help practitioners enhance their skills, improve their practice and solidify their ability to deliver a specific evidence-based model. Coaching includes skill modeling, shadowing (observing service provision), reflective supervision, collaborative discussion and immediate

feedback. Other strategies for coaching include site visits and telephone consultation from the program developers (or their staff members). New staff members implementing an evidence-based model need coaching to ensure model fidelity and high-quality services. Moreover, evidence suggests that all staff implementing evidence-based models benefit from ongoing coaching.[5]

- *Program evaluation technical assistance*: Evaluation drives the quality of implementation. Evaluations of model fidelity, consumer and staff satisfaction, and child/family outcomes are critical components of an agency evaluation system when delivering evidence-based programs.

- *Technical assistance for quality assurance*: Although collecting data is important, what one does with the data is also critical. Agencies implementing evidence-based programs will benefit from using program evaluation information to improve the quality of program delivery and outcomes for children and families. Agencies can enhance services through the establishment of feedback loops in which data are used by staff members and agency leadership to ensure high-quality programs. For example, classroom data may show that a new literacy program produces larger gains in reading skills when implemented in the morning rather than in the afternoon, or that improvements diminish when the program is administered to more than 10 students at a time. With a feedback loop, these concrete data would be used to alter program administration to optimize gains.

With all of these scaffolding components in place, proven programs stand the greatest chance of performing optimally and achieving expected outcomes on a large scale.

A Model from North Carolina: The Alliance for Evidence-Based Family Strengthening Programs

One model for implementing programs with fidelity has been developed by the Alliance for Evidence-Based Family Strengthening Programs

(the Alliance). This group is a unique, collaborative effort of public and private funders in North Carolina created to fund and implement select evidence-based programs for children and families. The Alliance adheres to several guiding principles:

- When communities invest wisely in proven programs that promote optimal child development, the next generation will pay back the investment as productive and responsible citizens.

- Communities and agencies implementing proven programs need access to ongoing training, technical assistance and program evaluation to deliver these programs successfully.

- The Alliance members can best advance the replication of proven programs by working together, across funding streams, to support and sustain programs that evidence shows produce the best outcomes for children and their families.

> With scaffolding components in place, proven programs stand the greatest chance of performing optimally and achieving expected outcomes on a large scale.

The Alliance is committed to increasing the number of evidence-based programs available to all North Carolina communities and to supporting the infrastructure necessary to ensure quality services for children and families. Programs collaboratively supported by Alliance members must:

- Meet documented needs demonstrated by North Carolina's children and families,
- Have evidence of effectiveness through randomized trials, and
- Have the existing infrastructure (or the willingness to create that infrastructure) necessary for successful replication in North Carolina.

The Alliance is staffed by the statewide nonprofit organization Prevent Child Abuse North Carolina and Duke University's Center for Child and Family Policy. Alliance members include private philanthropic funders (The Duke Endowment and the Kate B. Reynolds Charitable Trust); government entities (NC Division of Mental Health, Developmental Disabilities, and Substance Abuse Services, NC Division of Public Health, NC Division of Social Services, NC Children's Trust Fund, and the Governor's Crime Commission); and the quasi-governmental NC Partnership for Children. The Alliance's work is a mechanism to leverage resources for proven programs among multiple funders to ensure the best outcomes for children and families.

How the Alliance Works

Individual members of the Alliance continue to fund a broad range of social programs outside of their work in the Alliance. However, within the Alliance they have agreed to support a select few programs with the goal of ensuring a high degree of quality in program implementation and thus better outcomes for the children and families participating in the select programs.

The collaborative work of the Alliance is based on several key premises adopted by all members:

Alliance members support common intermediate outcomes across many different programs. Researchers, practitioners and advocates have long "siloed" programs into narrow subtypes of prevention: substance abuse prevention, child abuse prevention, adolescent pregnancy prevention, juvenile delinquency prevention and so forth. Nevertheless, many of the evidence-based programs designed to strengthen families result in positive intermediate outcomes (e.g., improved parent/child relationship) that have long-term impacts across multiple domains (e.g., reduced use of substances by youth, reduced child maltreatment, improved school readiness). Thus, a program such as the Incredible Years (IY) that focuses on preventing and treating conduct disorders is also effective in reducing families' risk for child maltreatment and promoting school readiness among preschoolers. IY meets the goals of multiple public and private agencies and can be

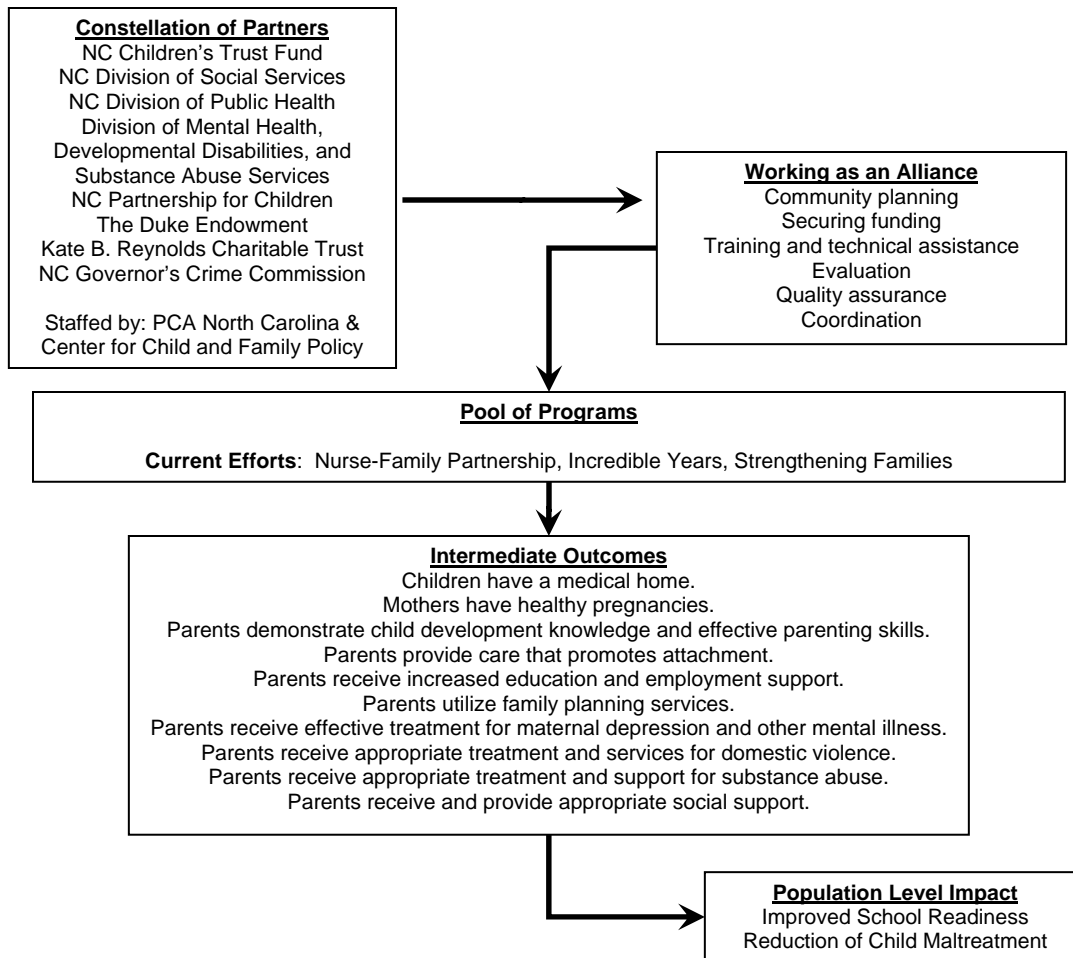supported collaboratively to maximize its impact in communities.

Figure 1 illustrates the Alliance's logic model, including the tasks of the Alliance, the funded programs, the intermediate outcomes and the targeted population-level outcomes.

Family strengthening programs selected for collaborative support by the Alliance receive scaffolding to ensure the most successful implementation of the programs in community-based settings. Key to the work of the Alliance is that selected programs include essential infrastructure that is consistent with the core drivers of successful implementation established in research. For program models that include infrastructure support from the "home" or national office of the program, Alliance funders provide support for communities to purchase these program services. Where national or "home" level infrastructure support does not exist, Alliance members are committed to creating that infrastructure at the state level in a unique collaborative model (described below) to provide coordinated scaffolding to local programs.

Collaboratively building implementation support for several select programs maximizes public and private dollars and will create efficiencies in program operations. Many evidence-based programs are already being funded by different public and private agencies across North Carolina. Although many of these programs have evidence of long-term cost-benefits to the public, they are often complex, more expensive to run and require greater expertise and capacity among staff members and agencies who are implementing the program.

Currently, individual funding agencies and/or local community-based agencies are negotiating and coordinating with national program developers for training, consultation and support of several evidence-based programs in North Carolina. Rather than having multiple North Carolina agencies working with national program developers to start and sustain a specific program (e.g., the Incredible Years), it is more efficient and effective to create a *collaborative portal of entry*

Figure 1: Alliance for Evidence-Based Family Strengthening Programs Logic Model

Constellation of Partners
NC Children's Trust Fund
NC Division of Social Services
NC Division of Public Health
Division of Mental Health,
Developmental Disabilities, and
Substance Abuse Services
NC Partnership for Children
The Duke Endowment
Kate B. Reynolds Charitable Trust
NC Governor's Crime Commission

Staffed by: PCA North Carolina &
Center for Child and Family Policy

Working as an Alliance
Community planning
Securing funding
Training and technical assistance
Evaluation
Quality assurance
Coordination

Pool of Programs

Current Efforts: Nurse-Family Partnership, Incredible Years, Strengthening Families

Intermediate Outcomes
Children have a medical home.
Mothers have healthy pregnancies.
Parents demonstrate child development knowledge and effective parenting skills.
Parents provide care that promotes attachment.
Parents receive increased education and employment support.
Parents utilize family planning services.
Parents receive effective treatment for maternal depression and other mental illness.
Parents receive appropriate treatment and services for domestic violence.
Parents receive appropriate treatment and support for substance abuse.
Parents receive and provide appropriate social support.

Population Level Impact
Improved School Readiness
Reduction of Child Maltreatment

for select programs that have strong evidence and are (or will be) used widely across the state. A collaborative portal of entry is beneficial for everyone: funders, community-based agencies, children and families served by the programs, and North Carolina taxpayers.

- A collaborative portal provides one point of access for all scaffolding, including information, training, coaching, program evaluation, quality assurance and technical assistance.

- Centralized staff provide this scaffolding to programs across the state, reducing replication

across systems. Instead of each agency creating fidelity tools, evaluation systems and coaching processes for their grantees, there is one set of these supports across all agencies.

- Centralization improves efficiency and allows for consistent, coordinated services and program monitoring.

- Local communities require fewer resources to successfully implement programs. Infrastructure necessary for quality program implementation (e.g., staff training, data collection and evaluation systems) is both more accessible and more affordable.

This collaboration and joint decision making regarding program implementation is a time-consuming and complex process. Partners must work within funding and policy mandates to create a shared set of implementation tools. Nevertheless, a collaborative portal of entry for specific evidence-based programs has the potential to save money and resources for the state and private funders alike, while improving program implementation. The cost savings will allow more agencies to provide quality, evidence-based services to ensure better program outcomes and, ultimately, improved well-being for children and families.

The Current Work of the Alliance

The Alliance is currently supporting two evidence-based programs that it hopes to bring to scale statewide and one program that has shown substantial promise:

- *Nurse-Family Partnership,* supporting highly trained nurses in visiting first-time, low-income mothers starting in pregnancy and continuing through their children's second birthdays;
- *Incredible Years Parent Training Programs,* providing group-based support to strengthen parenting competencies and foster parents' involvement in children's school experiences; and
- *Strengthening Families Program* (for which the Alliance offers limited implementation support), a parenting and family-building program for high-risk families that has been shown to reduce youth substance use and delinquency.

Extensive support for implementation of the Nurse-Family Partnership is provided by the program's national office to every implementing agency across the country (indeed, all local programs are required to participate in implementation support). Such extensive infrastructure is not available from the national Incredible Years office for local communities implementing the Incredible Years Basic Parent Training Program (IY). Thus, the Alliance has committed its resources to collaboratively supporting the development of such an infrastructure in North Carolina.

Building the Infrastructure for the Incredible Years

Two staff members, an IY statewide coordinator and an IY coach, are jointly funded by Alliance members and are housed at Prevent Child Abuse North Carolina. The statewide coordinator provides readiness assessment and support for local communities, marketing for the program and coordination of the program overall. The IY coach provides intensive clinical assistance to group facilitators to help them implement the program with quality and fidelity to program standards. Alliance funders contribute to the infrastructure of the Incredible Years by directly funding the staff positions at Prevent Child Abuse North Carolina, providing other supports (e.g., evaluation, trainings), and providing funding to local community-based agencies that are implementing IY and requiring that a "training and consultation fee" be paid back to the state IY office for these services.

Advantages of Working Collaboratively

Collaborative work among funders can be complex and requires considerable planning and consensus-building. Funders must build relationships with one another, develop a shared vision and goals, and pledge commitment to seeing the work through. Furthermore, staffing the infrastructure of the collaborative work requires resources. With these key ingredients, however, collaborative funding brings several significant advantages:

- There is an intentional planning process to identify priorities;
- Funders are focused on common outcomes.
- Development of a uniform vision strengthens the case for long-term sustainability;
- Funding and impact are maximized, not duplicated;
- Funders can compensate for one another's funding limitations and constraints;
- The public and private sectors are connected to fund the best interventions possible for children and families; and
- The implementation process is easier for communities (e.g., shared protocols, tools,

responding to announcements of funding availability).

Alliance Member Highlights and Achievements

This section highlights three Alliance members: a private foundation, a state agency and a statewide nonprofit.

The Duke Endowment

The Duke Endowment is one of the nation's largest private foundations. Its mission is to serve the people of North and South Carolina by supporting proven programs and strategies for higher education, health care, children's welfare and spiritual life. The Duke Endowment's Child Care Division provides funding to accredited organizations and other select nonprofit agencies that help children who do not have family support or who are at risk for losing such support. Funds go toward helping these children reach developmental milestones and prepare for successful transitions to adulthood. Since 1925, The Duke Endowment has provided approximately $2.4 billion in grants in North and South Carolina.

Within the past few years, the Child Care Division completed a strategic planning process to redefine its funding goals. This process resulted in a shift from funding services to funding outcomes. In other words, instead of simply funding programs that serve children in need, The Duke Endowment has moved to funding only programs with proven evidence that they change the lives of children. A key strategy is to support the replication, expansion or application of proven programs, defined as those with an established history and documented, positive results. "Proven programs" include the following:

- Effective practices, with evidence of multiple randomized controlled trials and replication in a real-world setting, as well as a sustained positive outcomes at least one year beyond the end of treatment; and
- Promising practices, with at least one control or comparison study that shows a greater likelihood of benefit than risk. (The Endowment funds these only if the program plans to include rigorous evaluation.)

The Endowment makes funding decisions by weighing the true costs and anticipated benefits of potential investments. Currently, the Child Care Division is funding seven proven programs, including Nurse-Family Partnership and Incredible Years.

When successful program models are lacking, The Duke Endowment invests in the development and testing of new approaches that may outperform existing practices. Typically, these begin with small-scale pilot studies to test key assumptions without major cost. If promising, larger evaluations may be funded.

With its shift in funding approach, The Duke Endowment has become committed to working collaboratively as part of the Alliance to disseminate evidence-based practices along with the necessary infrastructure to ensure optimal implementation.

The North Carolina Division of Social Services

The North Carolina Division of Social Services (NCDSS) provides oversight and support to the state's 100 county Departments of Social Services. Within the Division of Social Services, the Child Welfare Section oversees services to children and families with the goal of protecting children from child maltreatment. The Community-Based Team within the Child Welfare Section has the following mission: To provide children with safe, nurturing environments that promote their physical and emotional well-being by promoting protective factors and decreasing risk factors in families and communities. The Community-Based Team oversees several federal and state funding streams aimed at preventing maltreatment through the provision of family support/family strengthening programs. These funding streams include the federal Community-Based Child Abuse Prevention program and Promoting Safe and Stable Families program, as well as the state Family Resource Center program. Together, these funding streams represent approximately $2.2 million dollars that is distributed to local community-based agencies across North Carolina through a funding process that occurs every two years.

In a report released in 2005, the North Carolina Institute of Medicine Task Force on Child Abuse Prevention issued a set of recommendations for developing a statewide child maltreatment prevention system with the capacity to reduce the incidence of child abuse and neglect in North Carolina. These recommendations emphasized the importance of evidence-based practice to promote protective factors and decrease risk factors in families and communities.

In response to task force recommendations, NCDSS has made a fundamental shift to supporting the replication of proven programs in local communities. To facilitate that shift, NCDSS has changed policies and funding to support quality implementation of evidence-based programs. For example:

- For Family Support/Family Resource Program funding, 80 percent is required to go to evidence-based and promising programs and 20 percent to emerging practices;
- Regional training was provided to previously funded family resource centers to support a shift to evidence-based programming;
- NCDSS is investing in scaffolding for two evidence-based interventions for children and families, the Incredible Years and Strengthening Families, and provides funding for both of these programs at the local level; and
- NCDSS has invested in statewide training that provided community-based agencies with education on identifying and choosing evidence-based practices.

Communities and agencies implementing proven programs need access to ongoing training, technical assistance and program evaluation to deliver these programs successfully.

Prevent Child Abuse North Carolina

Prevent Child Abuse North Carolina (PCANC) is a statewide nonprofit dedicated to the prevention of child abuse and neglect in all its forms. PCANC provides implementation support for several evidence-based and promising family strengthening programs, conducts professional education and training activities, spearheads child abuse prevention public awareness efforts in the state, and serves as a leader for practice improvement efforts at the state and local levels.

While PCANC has long been interested in supporting best practices in local communities, its work to increase replication of evidence-based programs began in earnest when PCANC co-convened the statewide Task Force on Child Abuse Prevention with the North Carolina Institute of Medicine in 2005. That task force reviewed a number of evidence-based and promising family-strengthening programs and made specific recommendations to increase replication of those effective practices in the state.

To ensure implementation of the recommendations on evidence-based programs, PCANC and Duke University's Center for Child and Family Policy co-convened the Expert Workgroup on Evidence-Based Practice to explore strategies for increasing the replication of evidence-based family-strengthening programs. That Expert Work Group eventually became the Alliance for Evidence-Based Family Strengthening Programs. PCANC has helped to staff the collaborative meetings and work of the Alliance and has provided implementation support for a number of evidence-based and promising programs, including the Nurse-Family Partnership, the Incredible Years Parent Training Program, and the Circle of Parents Program. It also provides limited assistance for the Strengthening Families program.

36

The Alliance's Recommendations for Policymakers

The following recommendations stem from what the Alliance has learned from research on program implementation and from experience in implementing evidence-based programs.

1. Invest heavily in programs that have strong evidence of impact. Widespread use of evidence-based social programs will require changes in legislative and administrative policies.

2. A capacity-building approach is needed to shift policy and funding to evidence-based programs. Rather than putting out long lists of possible programs, pick a few programs that best fit the multiple needs of North Carolina. Fund them collaboratively and fully, with sufficient infrastructure support. Scant funding here and there for under-funded programs does not produce positive outcomes and in fact may harm families. (As the vaccine example at the beginning of the brief highlighted, watered-down programs are not effective.)

3. Shifting to evidence-based practices is a process that requires sustained attention and resources. It is critical to allow time for capacity building, careful program implementation and collection of quality outcome data. Implementing an evidence-based program likely takes at least two to four years before outcomes are evident.

4. Most successful programs will need skilled implementation support in an ongoing way. Both services and infrastructure are required in order for the programs to be successful. Implementation support ("scaffolding") will require additional resources, but the resulting improved outcomes translate into long-term cost savings. Policymakers are encouraged to provide funding and regulatory guidance, as needed, for the following implementation supports:
   a. Start-up costs associated with the program (e.g., community and agency assessment and planning, equipment, training);
   b. Support of the ongoing infrastructure for continued fidelity and sustainability (e.g., in-service training, ongoing consultation and coaching, fidelity measures); and
   c. Ongoing evaluation of program process and outcomes, along with quality assurance feedback loops to promote continual improvement of service quality and fit with community needs (evaluation is not an add-on but a fundamental piece of getting better outcomes and implementing evidence-based programs).

5. Success will best occur through collaborative efforts across multiple funding streams.

Many of these lessons can be applied across a broad range of policy domains, such as transportation, housing and economic development, in addition to social programs for children and families. With policy and programming built on the above recommendations, program recipients will benefit from improved services and greater well-being. Likewise, North Carolina citizens will benefit from efficient and effective use of resources, with long-term fiscal benefits surpassing costs.

[1] Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M. & Wallace, F. (2005). *Implementation research*: *A synthesis of the literature*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

[2] Joyce, B., & Showers, B. (2002). *Student achievement through staff development (3rd ed.)*. Alexandria, VA: Association for Supervision and Curriculum Development.

[3] Henggeler, S. W., Melton, G. B., Brondino, M. J., Scherer, D. G., & Hanley, J. H. (1997). Multisystemic therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology, 65*(5), 821-833.

[4] Gottfredson, D., Kumpfer, K., Polizzi-Fox, D., Wilson, D. Puryear, V., Beatty, P., & Vulmenay, M. (2006). The Strengthening Washington DC Families project: A randomized effectiveness trial of family-based prevention. *Prevention Science, 7*(1), 57-74.

[5] Schoenwald, S. K., Sheidow, A. J., & Letourneau, E. J. (2004). Toward effective quality assurance in evidence-based practice: Links between expert consultation, therapist fidelity, and child outcomes. *Journal of Clinical Child and Adolescent Psychology, 33*(1), 94-104.

Brief 5

The Impact and Cost-effectiveness of Taking Evidence-based Prevention to Scale: Lessons Learned from One State's Experience

Brian K. Bumbarger and Jenni Owen

Overview

Over the past decade, Pennsylvania has promoted programs that have clear evidence of effectiveness. The state has created a dedicated funding stream to encourage communities to adopt these programs aimed at preventing and reducing youth violence, delinquency and substance use. This brief draws heavily from the Pennsylvania experience, which is discussed in greater detail in the six research publications referenced at the end of the brief.

Since 1998, Pennsylvania's Commission on Crime and Delinquency (PCCD) has taken a policy position of promoting more efficient use of state funds by increasing reliance on evidence to guide programming. PCCD has used grants to encourage local communities to replicate specific evidence-based programs that have been proven effective in well-designed research studies. Under this initiative, the Commission's Office of Juvenile Justice and Delinquency Prevention has invested over $60 million to support nearly 200 implementations of effective programs in more than 100 Pennsylvania communities. The benefits have been enormous. What follows describes how this happened, the benefits of using research and analysis to drive public policy, and what observers of the Pennsylvania experience have learned about strategies for using analysis to improve public policy.

[Note: For purposes of this brief, *evidence-based program* has the same meaning as *evidence-based intervention*.]

This brief highlights five key issues related to evidence-based programs:

- The importance of a common understanding of the term *evidence-based program* and of how rigorous research complements good public policy;
- The importance of considering rigorous program evaluation as the beginning and not the completion of a broad policy and practice agenda to improve public safety and public health;
- The importance of, and challenges to, ensuring high-quality implementation of and fidelity to evidence-based programs as they move from research to real-world contexts;
- The importance of, and challenges to, the long-term sustainability of evidence-based programs; and
- The potential for large-scale adoption of evidence-based programs to yield not only positive outcomes but cost savings.

An understanding of these issues and strategies to address them will enhance policymakers' ability to adopt an evidence-based approach to policymaking and to the implementation of programs that stem from policymaking.

Evidence-Based Programs: What and Why?

The term *evidence-based* is becoming more and more common in policy and practice and has become synonymous with *effective*. However, the term is often used erroneously or without a clear understanding of what qualifies as being evidence-based. In setting policy to support an evidence-based agenda, it is important to have—and to communicate—a clear understanding of the term and its significance to efforts to promote public safety and public health.

In the area of youth violence and delinquency, the idea of evidence-based programs grew out of an effort by the University of Colorado's Center for the Study and Prevention of Violence (CSPV), funded by the federal Office of Juvenile Justice and Delinquency Prevention, to identify programs that had been shown to be effective in published research studies. CSPV's effort was called Blueprints for Violence Prevention and was one of the first efforts to apply specific, objective criteria when calling a program "effective." CSPV's criteria for evaluating a program's evidence of effectiveness included the following:

- An evaluation study with a strong research design;
- Evidence of significant prevention or deterrent effects;
- Replication of the positive findings in more than one study; and
- Effects that are sustained beyond immediate post-test.

Perhaps the most important of these criteria defining evidence-based programs is the strength of the study design. Randomized-controlled trials (RCT) are the gold standard of study designs. Because they use a control or comparison group and random assignment of test subjects, RCT studies represent an evaluation design that provides the strongest evidence and greatest confidence that positive results can be attributed to the program being evaluated. Without a control group, we cannot know that the changes we see are not simply part of a natural trend that would have happened regardless of the program. Further, without random assignment of the test subjects to

the intervention or control group, we cannot be sure that some other characteristic of the subjects (e.g., poverty level, intelligence, social support) is responsible for the positive changes we see.

Replication of positive findings in multiple studies is also important for ensuring that the program can generalize to different populations and contexts, and that the program's impact is not simply a result of the researcher's control over the study. Study design, replication and other key criteria are discussed in more detail in this report in the brief "Using Rigorous Evidence to Improve Government Effectiveness: An Introduction."

It is worth noting that since CSPV's original Blueprints project, different organizations and groups have developed a variety of criteria for judging the effectiveness of programs. These criteria vary slightly from one another, mainly as a function of the type of issue the group is interested in (violence, mental health, substance abuse, traffic safety, obesity, etc.), but there is considerable overlap, and all recognize the key importance of demonstrating effectiveness in a scientifically rigorous evaluation.

When comparing criteria promulgated by different groups, one should see them not as contradictory but as representing different points along a continuum of scientific rigor (and thus confidence). It is generally accepted that the CSPV criteria continue to represent one of the most rigorous standards of proof.

Programs that have demonstrated positive outcomes in rigorous and well-implemented evaluations can give policymakers the greatest possible confidence that they will be effective. Conversely, programs that do not have such empirical support can provide little confidence that they will have the intended impact and represent a greater risk of failure. Thus, evidence-based programs represent the safest bet when gambling with taxpayer dollars and the future of children and families.

As noted, Pennsylvania's Commission on Crime and Delinquency recognized the value of implementing programs that have clear evidence of

effectiveness and as a result has promoted the large-scale dissemination of effective programs aimed at preventing and reducing youth violence, delinquency and drug use.

Going Beyond a "List" to Achieve Broad Public Impact

PCCD's initiative to increase communities' adoption of evidence-based (i.e., "effective") programs addresses a major hurdle in prevention: the overwhelming majority of prevention and intervention efforts currently in place are *not* evidence-based and are not very effective. However, a policy agenda to achieve a broad public impact must go beyond simply providing communities with a list of effective programs and the funding to implement them. A number of issues must be addressed, and supports put in place, for these programs to realize the promise of improving outcomes at the community and state levels:

- Community readiness for evidence-based programming;
- Strategies for ensuring high-quality implementation of evidence-based programs; and
- Promoting sustainability of the programs beyond seed-grant funding.

Readiness: When policymakers fund evidence-based programs, are the communities ready?

As noted elsewhere in this brief and throughout the briefing report, the now-common call for evidence-based policy often emerges in legislation with policymakers requiring communities (or schools, county juvenile justice entities, or other recipients of public funding) to use resources only for programs that are evidence-based. While there may or may not be opposition to implementing evidence-based programs, willingness and desire are not a proxy for preparedness and optimal conditions.

Multiple factors may contribute to a community's readiness for evidence-based program implementation. Key among the factors is *context*.

In every context of planning for program implementation, policymakers and implementers will have increased success if they ask certain core questions early on to determine readiness for the evidence-based program:

- How strong is the evidence of the program's effectiveness?
- What are the other options?
- Does the evidence apply to our population?
- Can the program developer support our site?
- What will it take to sustain the program?

Beyond these core questions are other questions that can be tailored to the particular community:

- Do we need the program? (Does it address an identified risk factor?)
- Is it worth the investment, and can we afford it?
- Can we assemble the necessary resources and stakeholders?
- Will our community find it acceptable?
- How broad might the impact be?
- Do we know others who have used this program?

Stemming from each community's unique context, potential readiness challenges may include the following:

- A disconnect between the community's needs and specific funding opportunities (often called the "follow-the-funding" trap);
- Inconsistent motivations for adopting evidence-based programs either between policymakers and the community or within the community itself; and
- A lack of pre-implementation planning.

Evidence-based programs should be thought of as hardy seeds, but not magic beans. Even the best program will be ineffective in a chaotic and unstable school or community that lacks the basic resources and infrastructure. Sometimes these infrastructure issues must be addressed first, before funding is provided to implement new programs. This issue is especially relevant when resources are strategically targeted to the highest-need communities.

Match programs with community needs. An evidence-based program that is not aligned with a community's needs is likely to be a misdirected and wasted effort. The key to achieving the full potential of evidence-based programs is beginning with a community needs assessment that identifies the specific risk and protective factors that are causing poor outcomes for youth and families, and then selecting evidence-based programs that target those very risk factors. For example, implementing an evidence-based program to reduce the incidence of methamphetamine use will have little impact if the overwhelming drug of choice in a community is cocaine.

If North Carolina legislators decide to adopt an evidence-based policy philosophy, numerous local-level structures exist to support it. Juvenile Crime Prevention Councils, Smart Start Partnerships and workforce development boards could serve as positive readiness platforms for evidence-based program implementation.

Focus on risk factors. Delinquency and youth drug use arise from a complex set of individual, family and community circumstances. Thus the constellation of issues that drive delinquency in one community may be very different from those of another community. A good example is the difference between densely populated urban areas and more sparsely populated suburbs or rural communities. Each may have similar problems driven by a very different set of risk factors. By focusing on the risk factors rather than the outcomes, communities acknowledge their unique nature and can get at the root causes of the problems they are trying to prevent.

Pre-implementation planning is key. Research has found that pre-implementation planning is frequently lacking or nonexistent despite being identified as highly valuable. At least two reasons may explain this. First, there are often no funds allocated by policymakers for pre-planning activities or even requested for pre-planning in program budgets. Second, it is not uncommon that the communities most in need of prevention programs are the least well-equipped to plan for and implement them.

In Pennsylvania, PCCD addressed readiness using the Communities That Care risk-focused model for community mobilization. Before any community can access state funds to support evidence-based programs, it must form a community coalition and undertake a data-driven risk and resource assessment. Based on the data collected, each community creates a unique risk profile and prioritizes three to five risk and protective factors to address. The community then selects evidence-based programs that target those specific risk and protective factors. The resource assessment ensures that there are not existing programs or services already targeting these issues, and if there are, it examines whether the community has evidence about the impact of the existing programs. The additional benefit of this approach is that it enables all of the community's child-serving agencies and organizations to work toward a collective goal, creating synergy and economies of scale.

In a down economy, policymakers may be least likely to incorporate such readiness considerations into legislation and funding for evidence-based programs, but it may be the most important time to do so. Funding is too scarce; policymakers cannot afford to let communities fail due to readiness issues. PCCD's four-year funding model is an approach for North Carolina and other states to consider, as the cost-benefit discussion later in this brief suggests.

Quality implementation: We implemented most of the evidence-based program in our community, but…

The challenge of high-quality implementation and the issue of local adaptation remain the greatest barriers to achieving the promise of effective prevention. The best intentions for thorough and complete implementation of an evidence-based program may nonetheless result in partial implementation and program adaptations (intentional or unintentional) that may have a negative impact on outcomes. A significant body of research has shown that implementation quality and fidelity (i.e., doing the program exactly as prescribed) are strongly associated with better outcomes (although there remains some question as to precisely what counts as full implementation).

41

There is also considerable evidence that when these effective programs are implemented under natural (nonresearch) conditions, there is great variation in implementation quality.

Factors That Influence Implementation

Research on Pennsylvania's initiative by Pennsylvania State University's Prevention Research Center has identified five factors that may influence implementation in terms of program fidelity. These factors may also have an impact on program quality and adaptation, that is, the extent to which the implementation of an evidence-based program differs from the program as originally developed.

- *The implementer*: Does the implementer have a thorough understanding of the components of the evidence-based program and specifically of the program's logic model—that is, how the components and processes of the evidence-based program lead to the desired outcomes?

- *The implementing organization*: Are the staffing, resources and overall organizational capacity of the implementing organization ripe for doing the evidence-based program?

- *The program*: Are the tools of the program conducive to quality implementation? Tools include tangible items such as manuals and supplies as well as less tangible components such as training, technical assistance and professional development.

- *Recipients*: Can the program efficiently and successfully identify and recruit participants? Some program implementers face much greater challenges in this regard than others, for reasons ranging from communication to trust among community stakeholders and program representatives.

- *Context*: Is the evidence-based program being implemented in the context for which it was developed and evaluated? For instance, is a universal program meant to be directed at all students within a certain grade being delivered instead as a targeted intervention for a small group of high-risk youth?

One of the most common arguments against evidence-based programs is that the emphasis on strict fidelity to the program model does not allow for local knowledge and experience that could be used to strengthen the program by adapting it to better fit the community. While this is a compelling argument, research and experience in Pennsylvania have found it to be overstated. Researchers at Penn State studied the adaptations made by local implementers and found that they were seldom intentional attempts to alter the program to better suit the community. Further, the research showed that 80 percent of the adaptations made were in conflict with the program's underlying theory and were likely to reduce the chances that the program would be effective.

When considering program implementation and adaptation, it is important to distinguish between program changes that stem from innovation and those that are the result of program drift. Importantly, in addressing these challenges of "adaptation" there is often a disconnect between practitioners and researchers on the topic. Theoretically, if a community identifies a program characteristic that if modified would benefit program recipients, such an innovation would be welcome. However, if the modification has not been tested but is simply the result of implementers' beliefs that it would be positive, or if the change is simply in response to some implementation barrier, the result could be just the opposite. Program drift is rarely the result of pro-active, well-intentioned innovation. It is more likely the result of challenges to implementation such as greater comfort with former practices, staff turnover, relationships, lack of time or other resources or other institutional or community barriers.

Addressing Implementation Challenges: Making Programs More Effective

We know that the process of replicating evidence-based programs inevitably results in a range of program changes. We know that communities face assets and barriers that have an impact on implementation of evidence-based programs. The

goal is to create an infrastructure that promotes the highest-quality implementation, the least amount of variability between programs, and a level of skill and understanding among practitioners that empowers them to make thoughtful decisions about program adaptation and implementation. A program's efficacy (ability to produce positive outcomes when all the conditions are right) is less important than its effectiveness (ability to produce positive outcomes in natural conditions).

> The key takeaway for policymakers is to recognize that while legislating implementation of evidence-based programs is a critical step toward positive outcomes, the implementation process itself likely makes the difference between program success and failure.

There are several strategies that can enhance effectiveness through quality implementation:

- In replicating an evidence-based program, ensure that the implementer (local practitioner) is well trained in the underlying theory of the program, not simply the mechanics of delivery.

- Encourage or require implementers to assess implementation quality and fidelity as a part of their ongoing program assessment activities. It is as important to collect implementation data (through observations or self-reports) as it is to collect program impact data.

- Support ongoing technical assistance. As is the case with funds for pre-implementation planning, there is often a dearth of resources for this critical need. As part of its funding commitment, PCCD is requiring certain types of technical assistance when needed to maintain or improve program implementation.

The key takeaway for policymakers is to recognize that while legislating implementation of evidence-based programs is a critical step toward positive outcomes, the implementation process itself likely makes the difference between program success and failure. Gaps are likely between policy, implementation and delivery of programs. Policymakers can give or leverage support to mitigate potential implementation challenges and can promote an infrastructure of sufficient training and technical assistance to ensure that programs are delivered with the quality necessary to achieve significant positive impact.

Sustainability: We implemented the evidence-based program, but after a couple of years…

Research has identified effective evidence-based programs. Policymakers and other funders have funded them. Communities have implemented them. But identification and implementation of programs proven to have positive outcomes does not ensure sustainability of those programs over time. Many implemented programs with positive outcomes are never institutionalized. When a program fades, positive outcomes often fade with it. This is particularly true in areas such as juvenile delinquency and substance abuse. In these and many other areas of problem human behaviors, prevention strategies and interventions are rarely one-time, quick-fix solutions to short-term problems. On the contrary, the strategies typically need to be ongoing and long-term as are the challenges they are striving to address.

Addressing Sustainability Challenges

Viewing evidence-based programs as embedded within a larger community effort to change child and family outcomes over time represents a more realistic approach than the common method of providing short-term funding accompanied by grandiose expectations. To this end, PCCD's grants to support evidence-based programs are for four years (with an increasing match requirement in years three and four to promote sustainability). This four-year funding approach has proven invaluable as it often takes the better part of the first year to get a program off the ground and an additional year to "work out the bugs" and have the program fully functioning. Given this reality, funding such programs for only one year would literally be a waste of resources. It is highly unlikely that the programs would result in significant positive impacts in such a short time.

Again, the Prevention Research Center assessed the sustainability of PCCD funded sites and identified a number of strategies that encourage long-term sustainability:

- Ensure community and school support. If the community is not motivated, change is much less likely and sustainability almost impossible. Evidence-based programs that reported strong support from key community stakeholders and school administrators (in the case of school-based programs) were significantly more likely to sustain beyond PCCD grant funding.

- Address sustainability early and as part of the overall implementation process and plan. Interestingly, the research found that sites were able to accurately predict their sustainability five years in advance, and that programs that engaged in active sustainability planning early in the implementation process were more likely to sustain.

- Determine community readiness and develop it if necessary. This includes cultivating buy-in from key stakeholders, including but not limited to community leaders (or leaders of the setting where implementation will occur), program recipients, implementers and potential future funders. In addition, when developing buy-in, it is critical not to overlook the importance of the implementing organization itself. If key program staff are not at the table at the outset, both implementation and sustainability likely will suffer.

- Related to buy-in, collaboration among stakeholders is key, opening opportunities for sharing ownership and responsibility for program outcomes and success. The research in Pennsylvania found that program sites that were closely connected to a well-functioning community coalition had greater broad-based community support.

- Ensure that basic elements of organizational capacity are in place, including but not limited to accounting systems and adequate physical space. Even if motivation, buy-in and claims of readiness are strong, organizational capacity challenges could hinder sustainability.

As researchers have generated knowledge about some of these key factors for sustainability, PCCD has added strategies to its package of evidence-

based program components that reflect these factors. For example, communities are required to identify continuation funding sources during the application process and to provide match funding in the third and fourth years of PCCD funding. They also are offered technical assistance on sustainability planning. In addition, applications for funding must include the endorsement of a local community prevention coalition, and grantees are required to provide quarterly updates to the coalition on the program's progress and challenges.

> PCCD's four-year funding model is an approach for North Carolina and other states to consider, as the cost-benefit discussion later in this brief suggests.

Finally, sustainability is important not only for those directly involved with an individual community program, but also for the community more broadly. If a community experiences too many "one-shot deals," the readiness component discussed above may suffer and along with it the trust and collaboration between and among implementers and recipients that is necessary for program effectiveness.

While sustainability planning may make strategic sense, an understandable tension must be acknowledged with regard to balancing planning for sustainability before knowing with certainty whether a program works under natural conditions. Too often implementers scramble (often unsuccessfully) as initial funding ends to identify support for sustainability. Policymakers and implementers are therefore advised to build in support for sustainability up front. Although securing such funds is challenging in the absence of demonstrated positive impact, the existing evidence base for these programs provides a foundation of confidence in their potential to achieve positive outcomes. When coupled with a good structure for maintaining and measuring implementation quality, communities have a strong argument for approaching local stakeholders for program support.

Determining Whether Evidence-Based Programs are Effective, Cost-Effective and Cost-Beneficial

Even when the most rigorous evaluations show that a program is effective, high-quality implementation of the program is accomplished, and sustainability is maintained, there looms the question, Is it worth it?

From a public cost-benefit perspective, there are at least three types of effective programs:

- *Effective*: Some programs are effective but are too costly. There are ample examples of programs that could make a difference for youth with behavioral problems but are simply not feasible to implement at scale due to cost and other community or implementation constraints. Consider a program that provides costly one-on-one therapeutic supervision to a child with a behavioral mental health diagnosis. The intervention may prevent acting-out behavior during the school day, but the minimal impact does not justify the expense.

- *Cost-effective*: Cost-effective programs are both effective and nonburdensome from a cost standpoint. A cost assessment of these programs demonstrates that the programs are essentially a wash with regard to funds. They have positive outcomes for program recipients that offset the program costs. Consider a family therapy program that works intensively with adolescents who are on the verge of going into expensive out-of-home placement. Although costly, the program generates an almost immediate cost-savings offset by preventing placement.

- *Cost-beneficial*: Cost-beneficial programs are effective and represent a cost savings. In these cases the program not only offsets its costs with positive outcomes, but it actually represents a return on that investment by reducing larger community or societal costs beyond the immediate program outcomes. Consider a program for high-risk teen mothers that fosters strong parent-child bonds in early childhood, which later translate to higher graduation rates, increased employment and income (and

subsequent tax revenue), and reduced crime and drug treatment.

Effective (from a program impact standpoint) does not always equate to *cost*-effective. Likewise *cost*-effective does not directly equate to the greatest positive impact. Thus the challenge for policymakers is to weigh the effectiveness of a program against the potential costs and the potential benefits to reach the approach that maximizes impact while minimizing (current and future) taxpayer burden.

The Pennsylvania Experience

Using an approach developed by the Washington State Institute for Public Policy, Penn State's Prevention Research Center conducted a cost-benefit assessment of Pennsylvania's delinquency prevention efforts to determine the costs and benefits of that state's juvenile justice programs. The Pennsylvania analysis applied data from its own delinquency prevention programs, examining seven of the most prolific programs supported by PCCD, and found that the programs not only have positive outcomes but that the benefits they provide yield a measurable return on investment.

The PCCD case is not simply a "pay now or pay later" scenario. A more accurate characterization would be "pay now and pay significantly less later." What the analysis demonstrates is not only the cost savings from preventing, in this case, secure confinement for youth, but actual fiscal benefits ranging from improved employment (and therefore taxable income) to reduced burden on other public systems such as welfare, drug treatment, and social services.

Of the seven prevention programs examined in the Penn State study, the number of program recipients ranged across programs from six youth to 2,100 students and 410 families. The total potential economic benefit statewide was calculated for each program with findings ranging from a potential benefit (above and beyond the cost of the program) of $378,000 for a program serving 11 youth to a potential benefit of over $136 million for a program serving 109 families. The economic benefits were calculated based on the seven programs' significant likelihood to have positive

outcomes in terms of reductions in crime, substance abuse, and violence and/or an increase in test scores, employment opportunities and high school graduation rates. Although the Penn State study examined only seven programs within one particular state funding stream, the return on these seven programs alone was calculated to be $317 million, above and beyond recovering the actual costs of the programs.

> The PCCD case is not simply a "pay now or pay later" scenario. A more accurate characterization would be "pay now and pay significantly less later."

The outcome projections for these programs were not pie in the sky, best-case scenarios but were based on the actual levels of participant impact the programs had shown in previous studies, monetized and pro-rated based on the length of time it would take for a program to realize those outcomes (for elementary students to reach the age of criminal responsibility, for instance). This combination of proven programs yielding positive outcomes in natural conditions *and* reduced costs represents a compelling argument for policymakers.

With state prisons and county jails across the country operating at well over capacity, and the average state spending three times as much on corrections as on higher education, the implication for policymakers is clear. A significant state investment in prevention, which once was a tough sell and considered by many as liberal fluff, is now a well-informed and fiscally responsible approach to improving public safety.

Being able to point to effective programs that actually reduce societal costs over time is a windfall for policymakers. What better motivation could there be to require implementation of evidence-based programs for prevention? Yet there are ongoing challenges with regard to support for prevention. Prevention programs take the long view and thus are not always able to show immediate results. Although research has demonstrated that very small improvements measured immediately after program participation turn into significant long-term improvements, the

fact remains that the promise of prevention may not be realized within a single legislative cycle. PCCD's work is compelling, especially in such tight budget times; it has data to boast a combination of effective, evidence-based programs that also demonstrate cost-benefits. Even in good economic times there is an argument to be made that increasing investment in effective prevention programs could offset the burgeoning prison population in Pennsylvania, North Carolina, and beyond.

Summing It Up: Evidence Up, Youth Violence Down, Money Saved

Since 1998 the Pennsylvania Commission on Crime and Delinquency has invested over $60 million to aid communities in nearly 200 implementations of effective prevention programs across the state. In addition to the initial funding—allocated in four-year blocks—PCCD has supported implementation of the evidence-based prevention programs with significant proactive training and technical assistance to ensure these programs are delivered with the high quality necessary to achieve positive outcomes, and are sustained beyond initial state funding. As a result, research by Penn State's Prevention Research Center has shown that communities that adopt this approach have lower rates of delinquency and youth drug use, and that the state's investment yields significant economic savings in reduced systems utilization.

The Pennsylvania experience is still in progress. Research continues to assess the impact of the individual evidence-based programs as well as of the overall evidence-based policy approach that the state has adopted for juvenile delinquency, substance abuse and violence prevention programs. In part because Pennsylvania is approaching this as a statewide initiative rather than piecemeal, and has engaged in a partnership with researchers, opportunities have emerged to enhance the initiative based on the knowledge generated from studying the process of taking these programs to scale.

Stemming from this body of research and experience, PCCD has shown a willingness to recognize and act on needed enhancements to the

initiative, such as new requirements in the areas of implementation-quality oversight and sustainability planning. Further, based on the success of this approach over the past decade, the state has recently created the Resource Center for Evidence-based and Promising Programs and Practices. This effort expands the initiative across multiple state agencies by bringing in the state departments of Public Welfare (including mental health), Health (including substance abuse prevention and treatment), and Education.

What we can learn from the Pennsylvania experience regarding identification, implementation and sustainability of evidence-based programs has implications that go well beyond youth violence and delinquency. The programs' establishment of community partnerships, such as the required community-based prevention coalitions, could serve to facilitate the implementation of evidence-based programs in other domains (such as childhood obesity) within the same communities. Pennsylvania's approach is a telling case study, a working example of the potential and realized benefits of evidence-based programs. It is about the implementation of "model prevention

programs" within a supportive infrastructure and the fact that doing so is a process, not a one-time, short-term commitment.

One point is clear from Pennsylvania's experience. It behooves policymakers and practitioners to be good stewards of taxpayer funds by focusing their efforts on the implementation of programs that research has proven have a positive impact. Moreover, when budgets are tight, as is the case as of this writing, a commitment to thoughtful and strategic use of funds becomes even more necessary and expected.

The publications listed below (from which stem much of the material in this brief) offer a multidimensional view of the benefits, challenges and aspirations of undertaking an initiative to promote a statewide philosophy of using research evidence to guide policy decision making, funding and practice to promote positive youth development and strong families. Further, the publications expand on the opportunities identified here for policymakers, practitioners and researchers to play active and mutually supporting roles in advancing the positive impacts that implementation of evidence-based programs can have for individuals and communities.

Reading list:

Bumbarger, B. K., & Perkins, D. F. (2008). After randomized trials: Issues related to dissemination of evidence-based interventions. *Journal of Children's Services, 3*(2), 53-61.

Chilenski, S. M., Bumbarger, B. K., Kyler, S., & Greenberg Ph.D., M. T. (2007). *Reducing youth violence and delinquency in Pennsylvania*: *PCCD'S Research-based Programs Initiative*. State College, PA: The Pennsylvania State University.

Dariotis, J. K., Bumbarger, B. K., Duncan, L. G., & Greenberg Ph.D., M. T. (2008). How do implementation efforts relate to program adherence? Examining the role of organizational, implementer, and program factors. *Journal of Community Psychology, 36*(6), 744-760.

Jones, D., Bumbarger, B. K., Greenberg, M. T., Greenwood, P., & Kyler, S. (2008). *The Economic Return on PCCD's Investment in Research-based Programs*: *A cost-benefit assessment of delinquency prevention in Pennsylvania*. State College, PA: The Pennsylvania State University.

Kyler, S., Bumbarger, B. K., & Greenberg Ph.D., M. T. (2005). *Technical Assistance Fact Sheets*: *Evidence-based Programs.* State College, PA: Prevention Research Center for the Promotion of Human Development.

Tibbits, M., Bumbarger, B. K., Kyler, S., & Perkins, D. F. *Sustaining Interventions.* State College, PA: The Pennsylvania State University, Prevention Research Center. (Currently under review.)

Appendix I: Relevant Acronyms

CSPV	Center for the Study and Prevention of Violence, University of Colorado at Boulder
CTC	Communities That Care
EBP	Evidenced-based policy, evidence-based program or evidence-based practice
EPIS Center	Evidence-based Prevention and Intervention Support Center, Pennsylvania State University
GDL	Graduated Driver Licensing
ITT	Intention-to-treat analysis
IY	Incredible Years
NCDSS	North Carolina Division of Social Services
NFP	Nurse-Family Partnership
NIRN	National Implementation Research Network
NREPP	National Registry of Evidence-based Programs and Practices
PCANC	Prevent Child Abuse North Carolina
PCCD	Pennsylvania Commission on Crime and Delinquency
RCT	Randomized controlled trial
RFA	Request for applications
SAMHSA	Substance Abuse and Mental Health Services Administration

Appendix II: Glossary of Relevant Terms

This glossary includes definitions of frequently used terms related to evidence-based policy. Common terms in the fields of program evaluation, implementation and research are included, along with technical terms from the briefs in this report. This list has been developed for the 2009 Family Impact Seminar and is not comprehensive. An underlined word indicates that the term can be found elsewhere in the glossary.

Adaptation
A modification of an intervention to meet the needs of different people or settings, such as translating materials into other languages or adjusting content to make it more culturally appropriate for a particular population.

Adverse effect
Any harmful or unwanted change in a study group resulting from the use of an intervention.

Attrition
The loss of study participants during the course of the study due to voluntary dropout or other reasons. Higher rates of attrition can potentially threaten the validity of studies.

Baseline
The point in a study just before the intervention or treatment begins. The information gathered at baseline is used to measure change in targeted outcomes over the course of the study.

Best practice
Generally accepted as a successful intervention and believed to improve consumer outcomes. Determination of best practice relies on professionals to combine the research literature with their clinical experience and opinions.

Blinded research
Research that hides information on study-group status (treatment versus control) from participants and/or researchers. This method is used to prevent research outcomes from being influenced by either the placebo effect or observer bias. In a single-blind experiment, participants do not know whether they have been assigned to the treatment or control groups. In a double-blind experiment, both the participants and the researchers conducting the study are unaware of participants' group assignment. Only after the data are collected and/or analyzed do the group assignments become known.

Comorbid
The presence of one or more disorders (or diseases) in addition to a primary disease or disorder.

Comparison group
A group of individuals that serves as the basis for comparison when assessing the effects of an intervention on a treatment group. A comparison group typically receives some treatment other than "services as usual" and is therefore distinguished from a control group, though the terms are sometimes used interchangeably. To form a valid comparison, the composition and characteristics of the comparison group should resemble those of the treatment group as closely as possible. Ideally this is accomplished using random assignment to groups.

Confounding variables
In an experiment, any characteristic that differs between the experimental group and the <u>comparison group</u> at <u>baseline</u> (other than the <u>independent variable</u> under study). These characteristics or variables "confound" one's ability to explain the experimental results because they provide an alternative explanation for any observed differences in <u>outcome</u>. In assessing a classroom curriculum, for example, a confounding variable would exist if one group of students was taught by a highly experienced instructor, while the other group of students was taught by a less experienced instructor. Differences in student outcomes (e.g., grades) could be caused by the effects of the curriculum or by the variation in instructor experience.

Contamination / crossover
Occurs when individuals assigned to the <u>control group</u> inadvertently receive services or treatments intended for treatment group members. This biases study findings.

Control group
A group of individuals that serves as the basis of comparison when assessing the effects of an <u>intervention</u> on a treatment group. Depending upon the study design, a control group may receive no treatment, a "usual" or "standard" treatment ("treatment as usual"), or a placebo. To form a valid comparison, the composition and characteristics of the control group should resemble those of the treatment group as closely as possible. Ideally this is accomplished using <u>random assignment</u> to groups.

Correlational study
Studies used to look for relationships between variables. The three possible results of a correlational study are a positive correlation, a negative correlation and no correlation. The correlation coefficient is a measure of correlation strength and can range from –1.00 to +1.00.
- **Positive correlation:** Both variables increase or decrease at the same time. A correlation coefficient close to +1.00 indicates a strong positive correlation.
- **Negative correlation:** As the amount of one variable increases, the other decreases (and vice versa). A correlation coefficient close to -1.00 indicates a strong negative correlation.
- **No correlation:** A correlation coefficient close to 0 indicates no relationship between the two variables.

This type of study allows researchers to identify possible risk factors or <u>comorbid</u> conditions, but cannot prove causality.

Cost-benefit analysis
Analysis that weighs the total expected costs of an <u>intervention</u> against the total expected benefits of an intervention. Costs may include money, time and other resources spent to implement an intervention. Benefits may include reduced costs (both interpersonal and monetary) associated with short-term and long-term changes in participant <u>outcomes</u> (e.g., reduction in crime, reduced welfare use, increased graduation rates, etc.).

Cost-effectiveness
Economic analysis that compares the relative expenditure (costs) and <u>outcomes</u> (effects) of two or more courses of action to determine the most beneficial option.

Covariate
A secondary variable (not one of primary interest) that is possibly predictive of the <u>outcome</u> under study. For example, in a study on the effects of an <u>intervention</u> on reading skills, student age may be considered a covariate (because age and reading skills are related). A covariate can affect the relationship between the intervention and the outcome of primary interest.

Covariate analysis / analysis of covariance (ANCOVA)
Analysis that tests for intervention effects after removing the variance accounted for by other variables (e.g., age, sex). For example, in a study examining the effects of an intervention on reading skills, one may want to ensure that differences in student ages are not affecting the results. A covariate analysis including student age would adjust the intervention effect (reading skills, in this case) to a particular age, such as the mean age of all students.

Cross-sectional research
Observation of some subset of a population all at the same time, providing a snapshot of the current situation. Different age groups can be compared at a single point in time.

Culturally appropriate
Generally used to describe interventions or practices that show sensitivity to the differences among ethnic, racial and/or linguistic groups, as well as awareness of how individuals' cultural background, beliefs, traditions, socioeconomic status, history and other factors affect their needs and how they respond to the intervention.

Cultural competence
The knowledge and sensitivity necessary to tailor interventions and services to reflect the norms and culture of the target population and to avoid styles of behavior and communication that are inappropriate, marginalizing or offensive to that population. This trait is generally applied, with either positive or negative modifiers, to people or institutions.

Dependent variable
The observed phenomenon measured in an experiment, or the outcome (e.g., crime rate, graduation rate) that is affected during the experiment as a result of the independent variable (e.g., treatment type).

Effect size
A measure of the strength of the relationship between two variables, or the magnitude of the intervention effect. An effect size is calculated in a single study or across multiple studies to represent the size of a program's impact on any outcome that can be quantified.

Effectiveness
A measure of treatment impact in real-world settings. Effectiveness research is conducted with typical clients and routine practice conditions. Such research shows us how programs will work when replicated in nonresearch settings.

Efficacy
A measure of treatment impact under highly standardized research conditions. Efficacy research is highly controlled, which makes it easier to draw conclusions about program effects possible in ideal settings, but findings may not be generalizable to real-world settings.

Evidence
Scientific information regarding the effects of a well-defined treatment compared with a comparison group receiving no treatment, a placebo or an alternative treatment.

Evidence-based
Approaches to prevention or treatment that are based in theory and have undergone rigorous scientific evaluation. "Evidence-based" stands in contrast to approaches that are based on tradition, convention, belief or anecdotal evidence.

Evidence-based policy
Public policy informed by rigorously established objective evidence. Key aspects of evidence-based policymaking include:
- Evaluation of research findings to determine which programs have solid evidence of positive outcomes;
- Specific support, through funding and legislation, of evidence-based programs across policy realms, with careful attention to program implementation and ongoing outcomes; and
- Support of rigorous evaluation for innovative programs that are new or previously unstudied, to build the number of research-proven interventions.

Evidence-based practice
Skills, techniques and strategies that practitioners can use. Such practices describe core intervention components that have been shown to reliably produce desirable effects. They can be used individually or in combination to form more complex procedures or programs.

Evidence-based programs
Use of interventions for which systematic empirical research has provided evidence of significant effectiveness as treatments for specific problems. Top-tier evidence-based programs are those proven in well-designed and implemented randomized controlled trials, preferably conducted in natural community settings, to produce sizeable, sustained benefits to participants and/or society. Ideally, similar positive findings of such programs will have been observed by more than one evaluator and in more than one community. For the purposes of replication, programs should also have, at a minimum:
- Clear written guidelines for implementation (i.e. a manual or curriculum); and
- Mechanisms for monitoring intervention fidelity.

Experimental study design
A study design in which: (1) the intervention is compared with one or more control or comparison conditions, (2) subjects are randomly assigned to study conditions, and (3) data are collected at both pretest and posttest or at posttest only. The experimental study design is considered more rigorous than quasi-experimental or pre-experimental designs.

Expert-opinion approaches
Often called best practices, promising practices, or expert guidelines, these approaches rely on professionals to combine the research literature with their clinical experience and opinions. On one hand, expert-opinion approaches have the advantage of incorporating expertise other than research. On the other hand, expert-opinion approaches are subjective rather than based on rigorous evidence. As a result, expert-opinion approaches can be biased by cognitive errors, fads, current beliefs, idiosyncratic experiences, marketing and other factors that lead to invalid recommendations.

Fidelity
Occurs when implementers of an evidence-based program or intervention closely follow or adhere to the protocols and techniques that are defined as part of the intervention. Research on implementation of evidence-based programs shows that fidelity to core program elements is critical. Core elements (which may differ across programs) often include frequency/intensity of service delivery, the chronology of the service components, how services are delivered, staff education and professional credentials, training requirements, and substantive elements of the programmatic intervention.

Generalizability
The extent to which a study's results can be expected to occur with other people, settings, or conditions beyond those represented in the study. Rigorous study design (e.g., random assignment) and inclusion of a range of participant types will increase the generalizability of study findings.

Hierarchical modeling / hierarchical tests
Statistical analysis used with nested data (also known as multilevel analysis). For example, educational research may include data on pupils nested within classrooms nested within schools. Community research may include households nested within neighborhoods or counties. In studies where groups (e.g., classrooms, neighborhoods), rather than individuals, are randomized to treatment types, hierarchical modeling is the preferred method of data analysis, leading to more accurate conclusions.

Independent variable
Variable whose values are controlled or selected by the experimenter (e.g., treatment type).

Indicated
One of the three categories (Universal, Selective, Indicated) developed by the national Institute of Medicine to classify preventive interventions. Indicated prevention strategies focus on preventing the onset or development of problems in individuals who may be showing early signs of a disorder but are not yet meeting diagnostic levels of the disorder.

Intent-to-treat / intention-to-treat approach
An analysis based on the initial intentions for treatment as part of a study, not on the treatment eventually administered. All participants who begin the study are considered part of the trial and are analyzed based on their initial treatment assignment, whether or not they finish (or even start) the treatment. In other words, the outcomes of a participant who was randomly assigned to the active treatment group, but who failed to receive any of the treatment, will still be included in the treatment outcome data. This approach reduces the bias that may result when participants with lower motivation or more serious problems drop out of the treatment at higher rates.

Internal consistency
A measure of the correlation between different items on the same test. One aspect of test reliability, internal consistency examines whether items proposing to measure the same construct produce similar scores. Tests with adequate internal consistency (above .70 or .80) do a good job of measuring the intended outcome.

Internal validity
The degree to which a causal relationship between the treatment and the outcome can be confidently inferred from the experiment. Studies with high internal validity do a good job of controlling for alternative explanations and confounding variables, signifying that differences between the outcomes of the treatment and control groups are very likely due to the treatment itself. High internal validity is achieved through randomization, blinding and other rigorous experimental controls.

Intervention
A strategy or approach intended to prevent an undesirable outcome (preventive intervention), promote a desirable outcome (promotion intervention) or alter the course of an existing condition (treatment intervention).

Interrater reliability
The degree of agreement among individuals who rate the same item. If two or more individuals complete the same measure, the interrater reliability tells how much consensus there is between their ratings. If

raters do not agree, the scale may be defective or the raters may need to be retrained. Scores approaching +1.00 indicate high interrater reliability.

Logic model
A diagram, flow sheet or other type of visual schematic that communicates the logic behind a program, or its rationale. A logic model's purpose is to communicate the underlying theory, set of assumptions or hypotheses that program proponents have about why the program will work, or about why it is a good solution to an identified problem. What these schemata have in common is that they attempt to show the links in a chain of reasoning about "what causes what" in relationship to the desired <u>outcome</u> or goal. The desired outcome or goal is usually shown as the last link in the model.

Longitudinal research
A study involving repeated observation of the same subjects over long periods of time. Unlike <u>cross-sectional</u> studies, longitudinal studies track the same subjects over time, thus reducing the likelihood that observed differences are due to generational differences. This type of study may be used to examine trends or developmental pathways over time, or may be used to examine long-term <u>outcomes</u> of <u>interventions</u>.

Meta-analysis
A statistical tool that pools the results of multiple independent program evaluations and derives an overall program <u>effectiveness</u> rate.

Missing data
Data or information that researchers intended to collect during a study, but that was not collected or was collected incompletely. Missing data may occur, for example, when survey respondents do not answer all questions in a survey or when participants "drop out" of the study before it is complete. Large amounts of missing data can threaten the <u>validity</u> and <u>reliability</u> of a study.

Observer bias / observer effect
Changes in participant behavior as a result of being "watched" by the researcher. These changes may happen simply as a result of being observed, or researchers may unconsciously and subtly communicate their expectations to the participants, who alter their behavior to conform to these expectations. Observer bias is a significant threat to a research study's <u>internal validity</u> and is typically <u>controlled</u> for by using <u>blind experimenta</u>l designs.

Outcome
A change in behavior, physiology, attitudes or knowledge that can be quantified using standardized scales or assessment tools.

P-value
In research, this is the number by which one judges whether or not results are meaningful. To be considered <u>statistically significant</u>, an <u>outcome</u> effect should have a p-value of .05 or smaller (the smaller the p-value, the more likely it is that the results are meaningful). Technically, a p-value is the <u>probability</u> of obtaining a result at least as extreme as the observed result, when in fact there is no real effect. In a <u>randomized controlled trial</u>, a p-value of .05 means that there is only a 5 percent chance of obtaining the observed difference in treatment and <u>control group</u> outcomes if, in fact, the groups are equivalent.

Placebo effect
Phenomenon in which a person's beliefs about the effects of treatment with an inert substance or a sham therapy result in that treatment having the expected consequences upon that person's health. For

example, individuals who believe they are taking the real treatment (whether they are or not) show a stronger effect and those that think they are taking the placebo (whether they are or not) a lesser effect.

Population
The total set of subjects about which statistical inferences are to be drawn, often based on a random sample taken from the population.

Power
The probability that a statistical test will detect a treatment effect when, in fact, one exists. In other words, power is the likelihood of finding statistical significance when there is a real difference in outcomes between the treatment and control groups. Power is largely affected by sample size (i.e., number of study participants). In studies with low power, even a real treatment effect may not be identified.

Power analysis
An analysis to determine the power of a statistical test. *A priori* power analysis is conducted prior to the research study to determine an appropriate sample size to achieve adequate power (typically .80 or higher). *Post-hoc* power analysis is conducted after a study has been completed, and uses the obtained sample size and effect size to determine what the power was in the study.

Pre-experimental study design
A study design in which: (1) there are no (or poorly matched) control or comparison conditions, and/or (2) data are collected at pretest or posttest only. The pre-experimental study design provides less scientific rigor than experimental or quasi-experimental designs.

Pre-post study
A study that uses intervention recipients as their own control group by comparing pre-intervention scores on relevant measures with the scores received after the intervention is complete. This ensures group equivalence on key characteristics but fails to account for the passage of time or for other interventions/events that may have taken place concurrently.

Propensity scores
The probability of a unit (e.g., person, classroom, school) being assigned to a particular condition in a study given a set of known covariates (e.g., demographic characteristics). Propensity scores are used to reduce selection bias by equating groups based on these covariates.

Prospective study
A study in which the subjects are identified and then followed forward in time. A prospective study watches for outcomes, such as the development of a disease, during the study period and relates this to other factors such as suspected risk or protective factors.

Psychometrics
The field of study concerned with the theory and technique of educational and psychological measurement. The field is primarily concerned with the construction and validation of measurement instruments such as questionnaires and tests.

Quality assurance
Planned and systematic activities and processes used to check program fidelity and the quality of program implementation.

Quality improvement
Internally generated self-evaluation and improvement effort to ensure continuous progress toward meeting optimal standards. Optimal implementation of evidence-based programs includes the use of ongoing evaluation findings to inform continuous quality improvement.

Quasi-experimental study design
A study design in which: (1) the intervention is compared with one or more control or comparison conditions, (2) subjects are not randomly assigned to study conditions, and (3) data are collected at pretest and posttest or at posttest only. The quasi-experimental study design provides less scientific rigor than an experimental design and more scientific rigor than a pre-experimental design.

Random assignment
An experimental technique for assigning participants to different treatments. Once a study participant is identified, the treatment group for that participant is randomly selected, such as by the flip of a coin. Randomizing treatment assignment helps researchers ensure that treatment and control groups are equivalent at baseline. As a result, observed differences in outcomes between the two groups can be more confidently attributed to the treatment.

Randomized controlled trial (RCT)
A research design that provides the strongest, most reliable results about program effects, considered the gold standard in rigorous program evaluation. RCTs are characterized by: (1) comparison of at least two fundamentally different interventions (or one intervention versus "services as usual"), and (2) random assignment of recipients (individuals, groups, towns, etc.) to the different interventions in order to balance both the observed and unobserved differences among the groups (i.e., to ensure that the groups are equivalent). Other important aspects of RCTs include the use of reliable and valid instruments to assess outcomes, use of unbiased ("blinded") raters who are not aware of the intervention condition of the individual recipient, close monitoring of intervention conditions for fidelity and compliance, and analysis of results by standardized statistical procedures designed to maximize the statistical validity of conclusions.

Reliability (of a measure)
The consistency of a measure or test, either across time (test-retest reliability) or across raters (interrater reliability). Reliability does not imply validity—that is, a reliable measure is measuring something consistently, but is not necessarily measuring the right construct.

Replication
Repeating implementation (at a new site, with a new population, etc.) of a documented intervention.

Research-based intervention
An intervention based on researched theory, but not specifically evaluated as an intervention.

Retrospective study
A study that looks back in time to examine exposures to suspected risk or protection factors in relation to an outcome established at the start of a study. Errors due to confounding and bias are more common in retrospective studies than in prospective studies. If the outcome of interest is uncommon, however, the size of prospective investigation required to estimate relative risk is often too large to be feasible.

Sample
A group of subjects selected from a larger population, often for participation in a study. If researchers understand how likely it is that their sample is representative, they can draw meaningful conclusions from the sample about the population.

56

Sample size
The number of observations in a sample (e.g., the number of people in a study). All else being equal, a larger sample size gives a study more statistical <u>power</u> and increases the precision of <u>population</u> estimates (e.g., estimates of treatment effects if the <u>intervention</u> was delivered to the whole population).

Selective
One of the three categories (<u>Universal</u>, Selective, <u>Indicated</u>) developed by the national Institute of Medicine to classify preventive <u>interventions</u>. Selective prevention strategies focus on specific groups viewed as being at higher risk for negative <u>outcomes</u> because of highly correlated risk factors (e.g., children of parents with substance abuse problems).

Statistical significance
Used to define experimental results that are unlikely to have occurred by chance. To be considered statistically significant, an <u>outcome</u> effect should have a <u>p-value</u> of .05 or smaller (the lower the p-value, the more likely it is that the results are meaningful).

Target population
Those determined to be at greatest need of and highest priority for services and to whom the services are directed when the services are not <u>universal</u>.

Universal
One of the three categories (Universal, <u>Selective</u>, <u>Indicated</u>) developed by the national Institute of Medicine to classify preventive <u>interventions</u>. Universal prevention strategies address the entire <u>population</u> (national, local community, school, neighborhood) with efforts to prevent negative <u>outcomes</u>.

Validity of measure
The degree to which a test measures what it was designed to measure. Types of validity include:
- *Construct validity*: the degree to which inferences can legitimately be made from the measures in a study to the theoretical constructs on which those measures were based. Construct validity is demonstrated by establishing:
  - *Convergent validity*: the degree to which a measure is correlated with other measures that it is theoretically predicted to correlate with (corresponds with measures of similar constructs); and
  - *Discriminant validity*: the degree to which a measure is NOT correlated with other measures that it theoretically should not correlate with (that is, the researcher should be able to discriminate between dissimilar constructs).
- *Criterion validity*: The success of measures used for prediction or estimation. For instance, the correlation of measurement scores with actual behavioral <u>outcomes</u>.

Many of these definitions are taken in part or in their entirety from:

North Carolina Science to Service Consortium. (2004). *Bridging science and service: A plan to implement evidence-based practices for adults with mental illness in North Carolina's public mental health system.* Raleigh, NC: Governor's Institute on Alcohol and Substance Abuse. www.governorsinstitute.org/index.php?option=com_docman&task=cat_view&gid=23&itemid=65

Drake, R., Latimer, E., Leff, S., McHugo, G., & Burns, B. (2004). What is evidence? *Child and Adolescent Psychiatric Clinics of North America, 13*, 717-728. www.governorsinstitute.org/index.php?option=com_docman&task=doc_view&gid=33&Itemid=65

Jones, D., Bumbarger, B., Greenberg, M., Greenwood, P., & Kyler, S. (2008). *The economic return on PCCD's investment in research-based programs*. State College, PA: Pennsylvania State University, The Prevention Research Center for the Promotion of Human Development.
prevention.psu.edu/pubs/docs/PCCD_Report2.pdf

Schmitz, C., & Parsons, B. (1999). Everything you wanted to know about logic models but were afraid to ask. Boulder, CO: InSites.
www.insites.org/documents/logmod.pdf

Substance Abuse and Mental Health Services Administration (SAMHSA)
nrepp.samhsa.gov/help-glossary.asp

Wikipedia
www.wikipedia.org
Note: Throughout this glossary, Wikipedia was used as a resource only when the editors of this briefing report were certain of the accuracy of the information provided.

Appendix III: Resources for Evidence-based Policy and Programs

This list of organizations, publications and other resources provides guidance to policymakers, researchers and practitioners seeking additional information about evidence-based programs and policy. The list is not comprehensive, and the authors of this report do not necessarily support the views presented in the listed organizations' materials. The language used is summarized from the Web sites and documents referenced.

Evidence: Definitions and Criteria

Centers for Disease Control, Division of Violence Prevention. (In preparation). *Continuum of evidence of effectiveness.* **Atlanta, GA: Centers for Disease Control and Prevention.**
Will be available at: www.cdc.gov/ncipc/dvp/dvp.htm
The Evidence Project has developed a "Framework for Thinking about Evidence," which includes:
- *Best Available Research Evidence (Continuum of Evidence of Effectiveness)*: based on a continuum of the strength of evidence (from weak to strong) and rigor across various domains of interest (from well-supported to concerning/harmful).
- *Experiential/Colloquial/Tacit (Craft) Knowledge Evidence*: based on insight, understanding, skill, and expertise that is accumulated over time.
- *Contextual Evidence*: based on practical issues such as whether a strategy is useful, feasible to implement, and accepted by a particular community.

Drake, R., Latimer, E., Leff, S., McHugo, G., & Burns, B. (2004). What is evidence? *Child and Adolescent Psychiatric Clinics of North America, 13***, 717-728.**
www.governorsinstitute.org/index.php?option=com_docman&task=doc_view&gid=33&Itemid=65
Reviews definitions, criteria, and strategies related to scientific evidence. Discusses critical issues regarding the nature of treatment evidence. Offers suggestions for further consideration in the process of synthesizing evidence for clinicians.

SAMHSA Criteria for Defining Evidence-Based Practices
ebp.networkofcare.org/definitions/index.cfm?pageName=Criteria
Provides criteria for evidence-based practices that have been developed to guide program review. Discusses how much evidence is needed; the drawbacks of social science research (resources, ethics, real world vs. clinical trials, homogenous sample pool); and the creation of an evidence base for programs that are used in the field currently.

Society for Prevention Research. (2004). *Standards of excellence: Criteria for efficacy, effectiveness and dissemination***. Fairfax, VA.**
www.preventionresearch.org/StandardsofEvidencebook.pdf
Articulates a set of principles for identifying prevention programs and policies that are sufficiently empirically validated to merit being called "tested and efficacious."

Urban Institute. (2008). *Beyond ideology, politics, and guesswork: The case for evidence-based policy.* **Washington, DC.**
www.urban.org/UploadedPDF/901189_evidencebased.pdf
Describes evidence-based policy and provides examples from the areas of health insurance coverage, education, sentencing policy, and redress for housing discrimination that show how evidence informs good policy (and lack of evidence can invite bad).

Models for Evidence-Based Policymaking

The Alliance for Evidence-Based Family Strengthening Programs, Raleigh, NC
A collaborative network of public and private funders who support the replication of specific evidence-based programs for children and families across North Carolina. The Alliance is committed to funding programs that have strong track records of producing results for children, families, and communities, and to funding the needed infrastructure for quality implementation of those programs. See Brief 4 in this report for a discussion of the Alliance's work. For more information, contact Sarah Currier at scurrier@preventchildabusenc.org.

Briss, P., Zaza, S., Pappaioanou, M., et al. (2000). Developing an evidence-based guide to community preventative services-methods. *American Journal of Preventive Medicine*, *18*(1S), 35-43.
www.thecommunityguide.org/methods/methods-ajpm-developing-guide.pdf
Includes worksheets on the process for assessing the strength of a body of evidence and the relationship between the strength of evidence and the strength of recommendations. Provides an overview of the process to systematically review evidence and translate that evidence into recommendations.

Chilenski, S., Bumbarger, B., Kyler, S. & Greenberg, M. (2007). *Reducing youth violence and delinquency in Pennsylvania: PCCD's research-based initiatives program*. **State College, PA: Pennsylvania State University, The Prevention Research Center for the Promotion of Human Development.**
www.prevention.psu.edu/pubs/docs/PCCD_ReducingYouthViolence.pdf
Details Pennsylvania's approach to effectively addressing juvenile crime. Presents case studies of model prevention programs being implemented in communities throughout the state through Pennsylvania Commission on Crime and Delinquency (PCCD) funding, and describes the resulting positive outcomes.

Coalition for Evidence-Based Policy, Washington, DC
prod.ceg.rd.net/Programs/ProgramDetail.cfm?ItemNumber=9711
Seeks to increase government effectiveness through rigorous evidence about "what works." Works with top Congressional and Executive Branch policymakers on evidence-based reforms. Particularly relevant publications include:

> **Coalition for Evidence-Based Policy. (2002).** *Bringing evidence-driven progress to education: A recommended strategy for the U.S. Department of Education*. **Washington, DC: The Council for Excellence in Government.**
> www.evidencebasedpolicy.org/docs/Evid-based_educ_strategy_for_ED.pdf
> Seeks to bring rapid, evidence-driven progress to US elementary and secondary education. Sets out specific recommendations for consideration by Education Department leadership, as well as the broader policy community including Congress.

> **Coalition for Evidence-Based Policy. (2003).** *Bringing evidence-driven progress to crime and substance-abuse policy: A recommended Federal strategy*. **Washington, DC: The Council for Excellence in Government.**
> www.evidencebasedpolicy.org/docs/Bringing_Evidence_Driven_Progress_to_Crime_Subst_%20Abuse.pdf
> Explores how the federal government can most effectively use its resources to advance the development and effective use of rigorous evidence on what works in crime and substance abuse policy. Sets out specific recommendations for consideration by the participating agency officials and the broader policy community, including Congress.

George, P. & Blasé, K. (2008). *Financing evidence-based programs and practices: Changing systems to promote effective service*. **Tampa, FL: University of South Florida, National Implementation Research Network.**
www.fpg.unc.edu/~nirn/resources/publications/financing_paper_final_2008.pdf
Proposes that to achieve optimal results, EBPs require a new service and funding paradigm, including: funding streams and program designs that support the use of effective practices and services, innovative financing, and financing that promotes positive outcomes.

Invest in Kids, Denver, CO
www.iik.org/about/history.php
Partners with communities to improve the health and well-being of Colorado's children (prenatal to age 5), particularly those from low-income families, through advancing programs that work. To implement its mission statewide, Invest in Kids employs a three-part strategy:
1. Identify high quality, research-based programs;
2. Facilitate implementation of programs in communities throughout Colorado; and
3. Promote sustainability of programs.

National Association of State Mental Health Program Directors Research Institute, Inc.
ebp.networkofcare.org/state_init/index.cfm?pageName=special
Provides a state-by-state description of initiatives to promote the development and use of evidence-based practices.

North Carolina Science to Service Consortium. (2004). *Bridging science and service: A plan to implement evidence-based practices for adults with mental illness in North Carolina's public mental health system*. **Raleigh, NC: Governor's Institute on Alcohol and Substance Abuse.**
www.governorsinstitute.org/index.php?option=com_docman&task=cat_view&gid=23&itemid=65
Proposes a comprehensive implementation agenda for evidence-based practices for adults with mental illness in North Carolina. Includes a strategic action plan with specific goals, objectives, and strategies.

Oregon Office of Mental Health and Addiction Services
(2004). *Project plan to promote the adoption of evidence-based practices*
www.oregon.gov/DHS/mentalhealth/ebp/proj-plan-promote.pdf
(2006). *Progress report on the implementation of evidence-based practices*
www.oregon.gov/DHS/mentalhealth/ebp/report2jud-com.pdf
Salem, OR: Oregon Department of Human Services
The Oregon Legislature directed the Department of Human Services and four other state agencies to spend increasing shares of public dollars on evidence-based services, culminating in 75% by the 2009-11 biennium. These agencies are required to report to the Legislature, in each budget period, an increasing proportion of funds that support evidence-based practices. These reports outline the DHS plan and progress in meeting these mandates.

Spoth, R., Greenberg, M., Bierman, K., & Redmond, C. (2004). PROSPER community-university partnership model for public education systems: Capacity-building for evidence-based, competence-building prevention. *Prevention Science, 5*(1), 31-39.
Presents a model to guide capacity-building in state public education systems for
delivery of evidence-based family and youth interventions designed to bolster youth competencies, learning, and overall positive development. Summarizes positive results of its implementation over a 12-year period in an ongoing project.

Task Force on Evidence-Based Practice for Children and Adolescents. (2008). *Disseminating evidence-based practice for children and adolescents: A systems approach to enhancing care.* **Washington, DC: American Psychological Association.**
www.apa.org/pi/cyf/evidencerpt.pdf
Discusses the challenges involved in integrating science and practice and developing systems to assure that children receive effective treatment. States that effectively implemented evidence-based practices require a contextual base, collaborative foundation, and creative partnership among families, practitioners, and researchers.

Effective Implementation

Center for Mental Health Quality and Accountability (2004). *Results of a survey of state directors of adult and child mental health services on implementation of evidence-based practices.* **Alexandria, VA: National Association of State Mental Health Program Directors Research Institute, Inc.**
ebp.networkofcare.org/uploads/Lilly_Full_Report_3340664.pdf
Presents results of a 50-state survey conducted to: (1) obtain detailed descriptive information on state mental health agencies' policies, strategies, and mechanisms for implementing evidence-based practices in mental health service systems for adults and children; (2) identify models of successful implementation; (3) identify challenges, barriers, and facilitators that influence statewide implementation; and (4) identify needs related to current and future implementation.

Centers for the Application of Prevention Technologies
captus.samhsa.gov/home.cfm
Regional technical assistance providers for the Center for Substance Abuse Prevention (CSAP), an agency of SAMHSA. Assists states/jurisdictions and community-based organizations in the application of evidence-based substance abuse prevention programs, practices, and policies.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M. & Wallace, F. (2005). *Implementation research: A synthesis of the literature.* **Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).**
www.fpg.unc.edu/~nirn/resources/detail.cfm?resourceID=31
Describes the current state of the science of implementation, and identifies what it will take to transmit innovative programs and practices to mental health, social services, juvenile justice, education, early childhood education, employment services, and substance abuse prevention and treatment. Summarizes findings from the review of the research literature on implementation and proposes frameworks for understanding effective implementation processes.

Gorman-Smith, D. (2006). *How to successfully implement evidence-based social programs: A brief overview for policymakers and program providers.* **Washington, DC: Coalition for Evidence-Based Policy.**
www.evidencebasedpolicy.org/docs/How_to_successfully_implement_eb_progs-final.pdf
Advises policymakers and program providers on steps they can take to help ensure successful implementation of an evidence-based intervention, so as to achieve effects similar to those found in the research.

National Implementation Research Network, Chapel Hill, NC
www.fpg.unc.edu/~nirn
Seeks to close the gap between science and service by improving the implementation of evidence-based programs and practices. Provides information about *how* to implement evidence-based programs and what aspects of implementation and infrastructure are critical to success.

Penn State Prevention Research Center, State College, PA
www.prevention.psu.edu
Hosts the Evidence-Based Prevention and Intervention Support Center. Examines how communities can work together with families, schools, community groups (social service, youth groups, the faith community), and industry to promote healthy lifestyles for children, youth, and families. Provides policy-relevant information on best practices in prevention to federal, state, and local governments.

Schoenwald, S. K., Sheidow, A. J., & Letourneau, E. J. (2004). Toward effective quality assurance in evidence-based practice: Links between expert consultation, therapist fidelity, and child outcomes. *Journal of Clinical Child and Adolescent Psychology, 33*(1), 94-104.
Describes issues of creating demand for evidence-based practices, along with strategies for dissemination, implementation, coaching, and continuous quality improvement.

Stirman, S. W., Crits-Christoph, P., & DeRubeis, R. J. (2004). Achieving successful dissemination of empirically supported psychotherapies: A synthesis of dissemination theory. *Clinical Psychology: Science and Practice, 11*(4), 343-359.
Summarizes the key elements of three dissemination models and proposes an integrated plan that may be useful in promoting wider and more effective dissemination of empirically supported psychotherapies. Also discusses barriers to dissemination.

Identification of Evidence-Based Programs

Across Policy Realms

Campbell Library of Systematic Reviews, Campbell Collaboration
www.campbellcollaboration.org/campbell_library/index.php
Provides online access to systematic reviews of programs and interventions in the areas of education, criminal justice and social welfare.

Center for Mental Health Quality and Accountability. (2005). *Synthesis of reviews of children's evidence-based practices.* **Alexandria, VA: National Association of State Mental Health Program Directors Research Institute, Inc.**
ebp.networkofcare.org/uploads/Synthesis_of_Reviews_of_the_Research_on_Evidence_Based_and_Promising_Practices_9592994.pdf
Compilation of interventions or programs that have been evaluated or rigorously tested and found to have varying degrees of evidence as to their effectiveness.

The Cochrane Collaboration
www.cochrane.org/
Strives to improve healthcare decision-making through systematic reviews of research on the effects of healthcare interventions. The Cochrane Collaboration identifies the strongest studies addressing a given issue, helping researchers and policymakers separate reliable information in properly done studies from less reliable or rigorous information.

The Community Guide, Centers for Disease Control and Prevention
www.thecommunityguide.org
Offers evidence-based recommendations for programs and policies to promote population health. Topics include alcohol, motor vehicle safety, physical activity, tobacco, cancer, nutrition, pregnancy, vaccines, diabetes, obesity, sexual behavior, violence, mental health, oral health, social environment, and worksite health promotion.

Institute of Medicine of the National Academies, Washington, DC
www.iom.edu
Provides unbiased, evidence-based information concerning health and science policy to policymakers, professionals, leaders in every sector of society, and the public at large. Topics include aging, child health, diseases, education, environment, food and nutrition, global health, health sciences, healthcare and quality, mental health, military and veterans, minority health, public health and prevention, public policy, treatment, women's health, and the workplace.

Promising Practices Network
www.promisingpractices.net
Provides evidence-based information about what works to improve the lives of children, youth, and families. Features summaries of programs and practices proven to improve outcomes for children.

Washington State Institute for Public Policy, Olympia, WA
www.wsipp.wa.gov/topic.asp?cat=19&subcat=0&dteSlct=0
Summarizes evidence-based programs and cost-benefit analyses for interventions in the following areas: developmental disabilities, child welfare/child maltreatment prevention, children's mental health, criminal justice, juvenile justice, and K-12 education.

Social Programs that Work
www.evidencebasedprograms.org/
Summarizes the results of well-designed randomized controlled trials in social policy that have particularly important policy implications, because they show, for example, that an intervention has a major effect, or that a widely-used intervention has little or no effect. Organized into the following substantive areas: early childhood, K-12 education, youth development, crime/violence prevention, substance abuse prevention and treatment, mental health, employment and welfare, and international development.

World Bank's Poverty Impact Evaluations Database
www1.worldbank.org/prem/poverty/ie/db/evaluationdb.cfm
User-friendly database of evaluations of poverty reduction programs in developing countries, which offers an option to limit searches to randomized controlled trials.

Education

The Best Evidence Encyclopedia
www.bestevidence.org/
Established by Johns Hopkins University's Center for Data-Driven Reform in Education to provide reliable, unbiased information on high-quality evaluations of educational programs.

Institute of Education Sciences. (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide*. **Washington, DC: US Department of Education.**
www.evidencebasedpolicy.org/docs/Identifying_and_Implementing_Educational_Practices.pdf

Provides educational practitioners with user-friendly tools to distinguish practices supported by rigorous evidence from those that are not.

What Works Clearinghouse, Institute of Education Sciences
ies.ed.gov/ncee/wwc/
Established by the US Department of Education to provide educators, policymakers, and the public with a central, independent, and trusted source of scientific evidence of what works in education.

Mental Health / Substance Abuse

Collection of Evidence-Based Treatment Modalities for Children and Adolescents with Mental Health Treatment Needs, Virginia Commission on Youth
coy.state.va.us/Modalities/contents.htm
Summarizes the symptoms and etiologies for child and adolescent mental health concerns, and lists evidence-based treatment modalities and practices recognized as effective for treating these disorders.

Evidence-Based Programs, Oregon Department of Human Services
www.oregon.gov/DHS/mentalhealth/ebp/main.shtml#gen
Identifies proven practices in addictions and mental health services which have undergone independent review. Also includes reports on program implementation and fidelity.

National Registry of Evidenced-based Programs and Practices, SAMHSA
nrepp.samhsa.gov
Offers a searchable database of interventions for the prevention and treatment of mental and substance use disorders. Meant to help people, agencies, and organizations implement effective programs and practices in their communities.

Transportation and Urban Design

Badland, H., & Schofield, G. (2005). Transport, urban design, and physical activity: An evidence-based update. *Transportation Research Part D: Transport and Environment, 10*(3), 177-196.
Describes evidence on urban design factors as they relate to physical activity. Provides strategies in urban design and transportation that foster physical activity to improve population health outcomes.

Baker, S. P., Chen, L., & Li, G. (2007). Nationwide review of graduated driver licensing. Washington, DC: AAA Foundation for Traffic Safety.
aaafoundation.org/pdf/NationwideReviewOfGDL.pdf
Describes graduated driver licensing systems across the country, along with evidence for reduction in crash rates. Concludes that the most restrictive graduated driver licensing programs are associated with the highest reductions of fatal crashes and injury crashes involving 16-year-old drivers (reducing rates by 38% and 40%, respectively). See Brief 3 for a full discussion on North Carolina's graduated driver licensing program.

Violence Prevention / Child Welfare

Blueprints for Violence Prevention, **Center for the Study and Prevention of Violence**
www.colorado.edu/cspv/blueprints
Identifies outstanding violence and drug prevention programs that meet a high scientific standard of effectiveness. Serves as a resource for governments, foundations, businesses, and other organizations trying to make informed judgments about their investments in violence and drug prevention programs.

The California Evidence-Based Clearinghouse for Child Welfare
www.cachildwelfareclearinghouse.org
Provides child welfare professionals with easy access to vital information about selected child welfare related programs. Each program is reviewed and rated utilizing <u>a scientific rating</u> scale to determine the level of research evidence for the program. The Web site's format enables users to conduct literature searches, review extensive literature, and understand and critique research methodology.

Kauffman Best Practices Project. (2004). *Closing the quality chasm in child abuse treatment: Identifying and disseminating best practices*. **San Diego, CA: Children's Hospital – San Diego, Chadwick Center for Children and Families.**
colleges.musc.edu/ncvc/resources_prof/kauffman_final_report.pdf
Summarizes the most effective evidence-based treatments for child trauma victims and strategies for disseminating these "best practices" to front-line practitioners.

Cost-Benefit Analyses

Aos, S., Lee, S., & Miller, M. (2008). *Evidence-based programs to prevent children from entering and remaining in the child welfare system: Benefits and costs for Washington*. **Olympia, WA: Washington State Institute for Public Policy.**
www.wsipp.wa.gov/rptfiles/08-07-3901.pdf
Studies three basic questions: (1) Is there credible evidence that specific programs "work" to improve welfare outcomes? (2) If so, do benefits outweigh program costs? (3) What would be the total net gain to Washington if these evidence-based programs were implemented more widely?

Aos, S., Lieb, R., Mayfield, J., Miller, M., & Pennucci, A. (2004). *Benefits and costs of prevention and early intervention programs for youth*. **Olympia, WA: Washington State Institute for Public Policy.**
www.wsipp.wa.gov/rptfiles/04-07-3901.pdf
Concludes that some prevention and early intervention programs for youth, if well-implemented, can achieve significantly more benefits than costs. Taxpayers will be better off if investments are made in these successful research-based programs.

Aos, S., Miller, M., & Drake, E. (2006). *Evidence-based public policy options to reduce future prison construction, criminal justice costs, and crime rates*. **Olympia, WA: Washington State Institute for Public Policy.**
www.wsipp.wa.gov/rptfiles/06-10-1201.pdf
Analyzes 571 rigorous comparison-group evaluations of adult corrections, juvenile corrections, and prevention programs, most of which were conducted in the United States. Estimates the benefits and costs of many of these evidence-based options. Projects the degree to which alternative "portfolios" of these programs could affect future prison construction needs, criminal justice costs, and crime rates in Washington.

Aos, S., Miller, M., & Mayfield, J. (2007). *Benefits and costs of K–12 educational policies: Evidence-based effects of class size reductions and full-day kindergarten*. **Olympia, WA: Washington State Institute for Public Policy.**
www.wsipp.wa.gov/rptfiles/07-03-2201.pdf
Describes Washington's initial cost-benefit findings for class size reductions and full-day vs. half-day kindergarten.

Aos, S., Phipps, P., & Barnoski, R. (2004). *Washington's drug offender sentencing alternative: An evaluation of benefits and costs*. **Olympia, WA: Washington State Institute for Public Policy.**

www.wsipp.wa.gov/rptfiles/05-01-1901.pdf
Describes an outcome evaluation and cost-benefit analysis of one of the state's sentencing laws for drug-involved felony offenders: the 1999 amendments to the Drug Offender Sentencing Alternative (DOSA).

Barnoski, R. & Aos, S. (2003). *Washington State's drug courts for adult defendants: Outcome evaluation and cost-benefit analysis*. **Olympia, WA: Washington State Institute for Public Policy.**
www.wsipp.wa.gov/rptfiles/drugcourtMar2003.pdf
Describes drug courts and summarizes their effects on recidivism across more than 30 studies. Calculates cost-benefit ratio of this criminal justice option.

Jones, D., Bumbarger, B., Greenberg, M., Greenwood, P., & Kyler, S. (2008). *The economic return on PCCD's investment in research-based programs*. **State College, PA: Pennsylvania State University, The Prevention Research Center for the Promotion of Human Development.**
prevention.psu.edu/pubs/docs/PCCD_Report2.pdf
Examines the return-on-investment for seven research-based programs that are supported by the Pennsylvania Commission on Crime and Delinquency. Concludes that these programs not only pay for themselves, but represent a potential $317 million return to Pennsylvania in terms of reduced corrections costs, welfare and social services burden, drug and mental health treatment, and increased employment and tax revenue.

NCFIS Overview and History

The North Carolina Family Impact Seminar (NCFIS) is part of the Policy Institute for Family Impact Seminars network. The Institute was founded in 1999 at the University of Wisconsin-Madison/Extension and continues the family impact mission of the federal Family Impact Seminar, which operated from 1976 to 1998 in Washington, DC.

Since the start of state-level Family Impact Seminars in 1998, more than 20 states have convened impact seminars for their state legislators on a wide range of policy issues that impact children and families.

Duke University's Center for Child and Family Policy directs the NCFIS. The Center became the home site for NCFIS in 2004 and directed the first FIS in the state in 2005.

For more information on the North Carolina Family Impact Seminar series: www.childandfamilypolicy.duke.edu/familyimpact.

For more information on the Policy Institute for Family Impact Seminars: www.familyimpactseminars.org.

Previous North Carolina Family Impact Seminars

2005: Medicaid Cost Containment Strategies in North Carolina and Other States
www.childandfamilypolicy.duke.edu/familyimpact/2005.html

2006: Children's Mental Health: Strategies for Providing High-quality and Cost-effective Care
www.childandfamilypolicy.duke.edu/familyimpact/2006.html

2007: Juvenile or Adult? Adolescent Offenders and the Line Between the Juvenile and Criminal Justice Systems
www.childandfamilypolicy.duke.edu/familyimpact/2007.html

2008: Dropout Prevention: Strategies for Improving High School Graduation Rates
www.pubpol.duke.edu/centers/child/familyimpact/2008.html

Acknowledgments

Bridging the gap between research and public
policy to improve the lives of children and families

Center for Child and Family Policy
Duke University
Box 90545
Durham NC 27708-0545
919.613.9303
www.childandfamilypolicy.duke.edu