Brief 1

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.

This brief was prepared in conjunction with a presentation delivered by Jon Baron at the 2009 North Carolina Family Impact Seminar, "Evidence-based Policy: Strategies for Improving Outcomes and Accountability."
Jon Baron, JD, MPA, is the executive director of the Coalition for Evidence-Based Policy in Washington, DC. Katie Rosanbalm, PhD, is a Research Scholar at the Center for Child and Family Policy, Duke University.

The key aspects of evidence-based policymaking include:

- The 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
- The support of rigorous evaluation for innovative programs that are new and/or previously unstudied, in order to build the number of research-proven interventions. Using pilot programs with requirements for clear results of effectiveness before widespread replication minimizes spending on suboptimal interventions.

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.

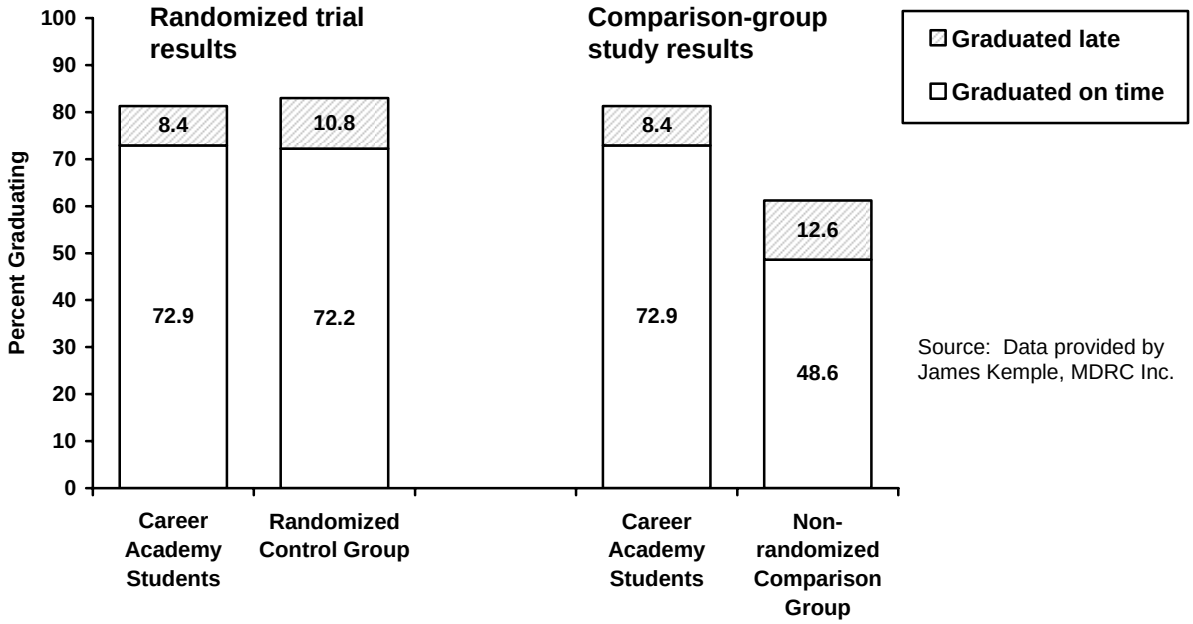
In these studies, statistical procedures (such as propensity scores or covariate analyses—see glossary for definitions of these terms) can be used in an attempt to "control for" group differences. Findings cannot always be trusted with a high level of certainty, however, as unobserved group differences 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



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

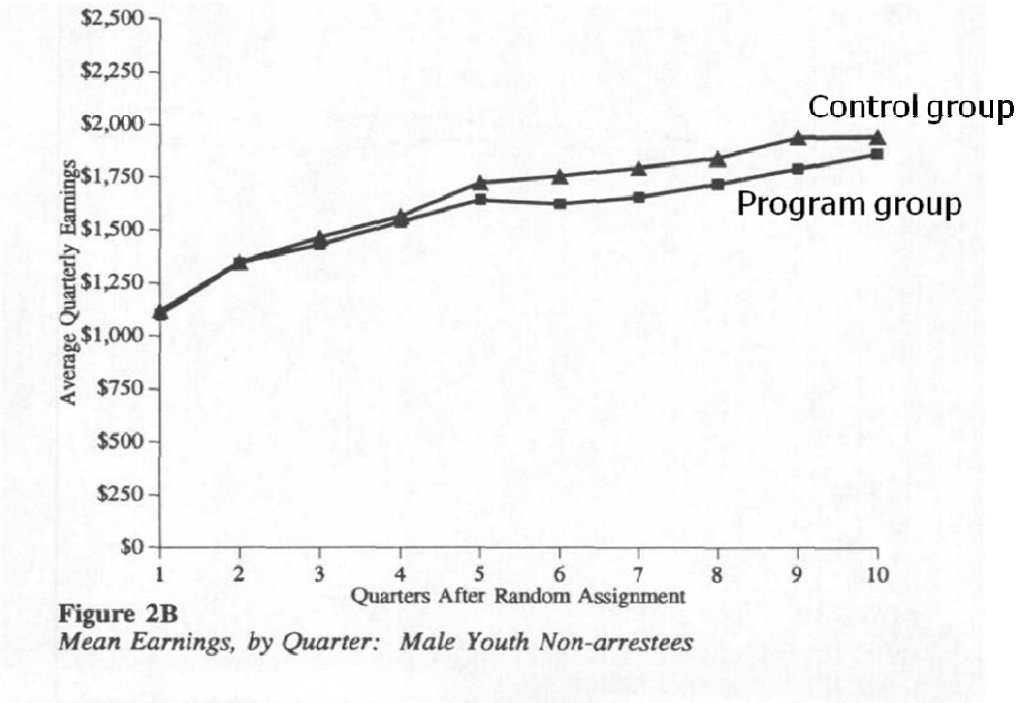
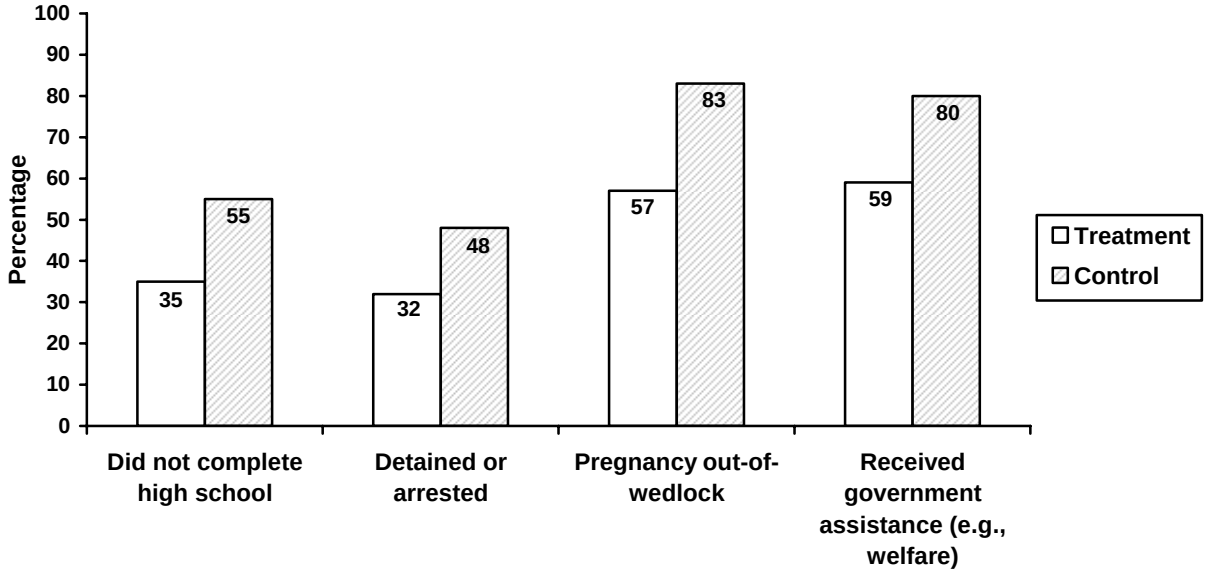


Figure 2B
Mean Earnings, by Quarter: Male Youth Non-arrestees

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*:

- 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.