Summing Up the Evidence from Experiments: Results from Gold Standard Reviews

Remarks Given by

Elizabeth Hair, Ph.D.
Senior Research Associate

Prepared by
Kristin Anderson Moore, Ph.D.
President and Senior Scholar

Child TRENDS

presented at

The Smith Richardson Foundation Meeting on New Findings from the C2 Testbed Project

May 28, 2003
The Inn at Perry Cabin
St. Michaels, MD

Summing Up the Evidence from Experiments: Results from Gold Standard Reviews

Remarks by Elizabeth Hair, Ph.D., Senior Research Associate, Child Trends, at the Smith Richardson Foundation Meeting on New Findings from the C2 Testbed Project on May 28, 2003. Prepared by Kristin Moore, Ph.D. and Elizabeth Hair, Ph.D.

Like a number of other research organizations, Child Trends staff have had the audacity to think that research might be useful to policy makers and program developers. It wasn't always clear, however, that policy makers and program developers felt the same way.

For a variety of reasons, research has played less of a role in public policy than many researchers have thought appropriate. There are probably a number of reasons for this.

One reason is rock solid and intrinsic to the nature of research. Specifically, this is the cumulative and changing nature of research, especially research on human behavior. The public would prefer that knowledge be simple and unchanging (like the boiling point of water), and that recommendations be simple and unchanging as well.

Of course, most science isn't simple and unchanging, especially the social sciences. As society changes, the phenomena that are studied change as well. And human behavior is extremely complex. Therefore, we social scientists start off at a disadvantage. Our knowledge base is only in the process of accumulating, and we are not likely to identify very many immutable truths.

Another factor that has undermined the role of research in public policy is that at least some researchers have gone beyond the data and shared their own opinions. In fact, it is part of "the game" for researchers to exaggerate in proposals in order to get funded.

Six weeks of sex education will stop teen pregnancy.
Driver's education will stop teens from drinking and driving.
Kids can be "scared straight" and will give up crime.

Of course these promises cannot be fulfilled, and when interventions that are highly touted don't work, this undermines the credibility of all researchers.

Another reason that research doesn't have much influence on programs or policies is that, to be frank, a lot of research is pretty weak and has little to offer. For example, cross sectional analyses that don't control for important confounding factors are often touted by advocates as holding important policy significance.

Eating dinner together provides a good example. Eating dinner together is correlated with better outcomes for kids, but it may in fact be time together and parent-child communication and family togetherness that really underlies this correlation. Eating dinner together may be a good solution; but, for parents who work the night shift or attend conferences away from home, there may be other ways to reach the goal as well as eating dinner together.

Of course, policy makers and program developers can't wait until researchers complete their final multivariate models with full controls for endogeneity and self-selection. We as researchers have not, however, been as careful as we might be in providing the caveats around the certainty of what we know. It is one thing to identify a construct like eating dinner together as a descriptive indicator; and quite another to make it a causal force that needs to be implemented in a treatment model.

Another reason is that researchers don't communicate well with people who could potentially use their information. We have all been at professional conferences where researchers slap up transparencies covered with coefficients that are unreadable even by other researchers. Policy makers and program developers, for the most part, can't and don't read academic journals with complex statistical models. They need good science that is simplified, but not so simplified as to be misleading.

This is the quandary that Child Trends faced when we were approached by the John L. and James S. Knight Foundation and the Edna McConnell Clark Foundation to help their staff think about program approaches to fund. How could research be useful?

To pick up another related but distinct strand of this history, it should be noted that Child Trends was heavily involved in the experimental evaluations of welfare reform during the 1990s, working with MDRC and with Howard Rolston in the Administration for Children and Families, and later with Mathematica and Abt as well. We also worked closely with the National Campaign to Prevent Teen Pregnancy, in developing the reports *No Easy Answers* and *Emerging Answers*.

The need for experimental evaluations, which are expensive and tricky to implement and carry out in the real world, was a topic of considerable debate throughout the 1990s (as everyone in this room knows). What works to help welfare recipients? What works to prevent teen pregnancy?

Are experiments necessary to answer these questions?

Work by Dan Friedlander and colleagues and by Rebecca Maynard and colleagues, and a very influential paper by Hollister and Hill made a very strong case that, yes, experimental studies are a breed apart from all other kinds of research. When well-conceived and well-implemented, experimental studies are indeed the "gold standard."

If so, then what is the point of other kinds of research? Why do we bother with basic research?

And what do we make of the wisdom of practitioners? If policy makers and program developers are a target audience for research, what are we accomplishing when we totally ignore their accumulated wisdom and their opinions?

In order to address these complicated issues, we ended up developing a new product: the *What Works* tables.

As the invitation to this conference notes, we faced the "hard choices between efficiency and getting to the truth."

It should be noted that Child Trends was not funded to do methodological work, or to advance discussions of how to sift evaluation evidence. We were funded to do a number of literature reviews, and we chose, based on our experience and our read of the methodological literature and working jointly with the Knight and Clark Foundations, to give more weight to experimental studies.

We literally put the experimental studies in a separate section and in separate columns of the *What Works* tables. (See http://www.childtrends.org/whatworks_intro.asp)

This work can be best illustrated with the adolescent development tables requested by the Knight Foundation. They needed information on what works to address seven adolescent outcomes, ranging from reproductive health to mental health to education to social relationships, so the Knight *What Works* tables are organized according to the studies that examined a given adolescent outcome. We reviewed over 1,100 studies, including both experimental, quasi-experimental and non-experimental studies in our review. In the review of experimental studies, we included experimental studies that had been through peer review or that were official government reports.

That's all. We did not select experimental studies based on sample size or other criteria. We treated the gold standard truly as a gold standard and made it the central organizing focus.

Then we did another very simple thing. We divided experimental studies that had positive impacts, negative impacts, and mixed impacts. The first column reports on "what works," that is studies that had positive impacts on the adolescent outcome in question. The second column reports "what doesn't work," that is, studies that had no impact or a negative impact on the adolescent outcome in question. The third column reports "mixed findings."

We would acknowledge that the "mixed findings" category puts studies that report subgroup impacts at a disadvantage. That was one of the hard choices that we made. We couldn't put them in the "What Works" column and we couldn't put them in the "What Doesn't Work" column, so we created a separate column to efficiently report these more complex findings.

Next we did something that has turned out spectacularly well. We created a final column called "Best Bets," to allow us to report findings – or promising practices – from non-experimental studies, including quasi-experimental studies and multivariate, longitudinal research, and the wisdom from the practice field.

We did this for two reasons. One reason is that there are so few experimental studies at this point in time that it is difficult to be useful if you limit your knowledge base solely to experimental studies. The second reason is that, while there is a lot of lousy research that had to be thrown out, there is a lot of good multivariate, longitudinal research that can inform policy and practice; and there are a lot of smart people working in programs and policy fields who have good ideas that deserve consideration.

Nevertheless, this information represents a different type of knowledge than the knowledge that results from experimental studies. So we separated them and labeled the columns clearly. Experimental. Non-experimental. We labeled them, and we included them both.

It turns out that this had some very fortuitous consequences. We have found that policy makers and program providers appreciated and felt respected by having their ideas represented. They don't mind having them listed in a "Best Bets" column. Also, policy makers and program providers don't want to be limited to the ideas that have been tested experimentally. Again, they don't mind having these ideas listed as "Best Bets." In fact, they seem to understand and even appreciate the difference.

For the Clark Foundation, we organized the *What Works* tables around program approaches. For example, we did a review of mentoring approaches and considered what works, what doesn't work, mixed results, and best bets for mentoring. Thus, while the literature review had a different goal, the *What Works* tables are organized similarly.

What did we find? Briefly, I want to highlight some of the findings from the Knight adolescent review. It is summarized in our "American Teens" report and the Research Brief, *Building a Better Teen-ager: A Summary of 'What Works' in Adolescent Development,* available at <http://www.childtrends.org/PDF/K7Brief.pdf>.

Let me suggest eight "take-away" points based on this review of 1,100 experimental and non-experimental studies.

1. Parent-child relationships are vital to adolescent well-being.

In other words, parents matter. Non-experimental basic research studies consistently confirm that teens who have warm, involved and satisfying relationships with their parents are more likely to do well in school, to have better social skills, and to have lower rates of risky sexual behavior than other teens. [It's hard to do an random assignment study of parents; but parents are too important to ignore just because we don't have experimental evidence.]

2. Peer influences are important and can be positive.

Again, non-experimental evidence suggests that adolescents often influence each other positively, by either modeling behaviors or by pressuring each other to behave in certain ways or to adopt certain attitudes and goals. Several experimental studies also indicate that interventions can improve peer interactions.

3. Siblings, teachers, and other adults and mentors can provide additional support.

Brothers and sisters can act as models for positive behaviors, such as physical activity and avoidance of drug use. Sibling relationships are also a good training ground for conflict resolution and negotiation skills.

Experimental studies further indicate that mentors can have a positive impact on adolescent outcomes. Non-experimental analyses suggest that mentors can offer guidance, friendship and assistance, and serve as positive role models, is positive relationships are developed and sustained over a period of time.

4. Adolescent behaviors often cluster.

Non-experimental research confirms the common sense observation that teens who show one positive or negative characteristic are likely to have other positive or negative characteristics. This clustering suggests that programs and policies can bring about change in a number of outcomes, even if their primary focus is on only one or two areas. Clustering also can alert parents and service providers that if an adolescent has one problem, he or she may have additional problems as well.

5. Teens should be viewed as whole people, more than just students, patients, or delinquents.

Taking a more complete and balanced approach to youth development requires an awareness of the many factors associated with adolescent well-being. For example, schools, communities, family income, the media, and public policies, all have implications for adolescent development. But many existing programs only target one or two aspects of youth (such as employment or substance abuse) and do not provide a comprehensive set of services that address the whole person.

6. Engage young people.

Experimental evaluations have shown repeatedly that didactic lectures do not change adolescent behavior, whether they target smoking, drugs, gun violence or teen pregnancy. On the other hand, many experimental evaluations find that adolescents who take part in programs that build relationships, truly involve teens, and provide well-implemented and structured activities tend to have lower rates of pregnancy and drug, alcohol and tobacco use.

7. It helps to start early and sustain the effort.

The problems that start in childhood often continue through adolescence and into adulthood. Intensive preschool programs such as the Abecedarian program confirm that such an early intervention affects not educational outcomes but also teen childbearing years later.

8. Think positively about teens.

Polls show that many adults see teens as walking billboards for problems. Thus, many programs directed at youth focus almost exclusively on preventing problems from occurring. However, an accumulating body of experimental and non-experimental research suggests that taking a positive approach – promoting skills and assets among teens -- may be a more effective way to avoid negative outcomes and help teens realize their potential.

In sum, we think that both experimental and non-experimental research have a lot to offer to policy makers. Our conclusions draw on both, and we have tried to share these conclusions in an efficient and user-friendly format.

It was not our goal, nor the goal of our funders, to undertake a methodological study. We just wanted to be useful, and we hope that sharing our thinking – and our findings – helps to further advance the discussion.