

In Support of Disciplined Passion

David Olds

Published online: 14 April 2009
© Springer Science + Business Media B.V. 2009

Abstract Manuel Eisner (2009) has posed the cynic's hypothesis to help explain the slippage found between developer-led investigations of preventive interventions and those led by independent evaluators, and he has called for more research on developer self-interest as a possible explanation for this pattern of results. The issue Eisner raises has important implications for the integrity of science and for policy making, but he gives insufficient attention to the challenges of intervention development and effective replication across populations, communities, organizations, and service providers. The challenge he poses needs to be addressed by higher standards for reporting trials, better peer review, improved investigator training, and rigorous, collegial support of those who choose to enter this challenging field, lest we dissuade young investigators from seeking careers that combine their commitment to solve social problems with their commitment to rigorously evaluate their efforts.

Keywords Bias · Experimenter · Integrity · Independent evaluation · Intervention development · Peer review · Replication

The well-known slippage between results derived from efficacy trials (conducted under relatively optimal conditions) and effectiveness trials (conducted under more real-life conditions) poses a significant challenge for evidence-based policy and practice. Reflecting on Petrosino and Soydan's (2005) finding that large effects on recidivism are reported for crime-reduction interventions tested by developers and no effects in replications examined by independent evaluators, Manuel Eisner (Eisner 2009) poses the cynic's hypothesis: that the difference observed is attributable to developers' self-interest. Eisner presents compelling findings from a variety of sources to support his hypothesis that developers' self-serving cognitive biases or desires to line their pockets or build their programs of research can lead the

D. Olds (✉)
Prevention Research Center for Family and Child Health, University of Colorado,
13121 East 17th Avenue, Room 5317, Aurora, CO 80045, USA
e-mail: David.Olds@ucdenver.edu

developer–investigator to compromise the conduct of trials at every step of the way. These are serious issues that need to be addressed thoroughly, because, as Eisner notes, public trust in the scientific enterprise and appropriate allocations of scarce public resources depend upon the validity of our estimates of intervention impact. He calls for more research to estimate the degree of bias produced by program developers' evaluating their own programs. While studying conflict of interest may ferret out the degree to which investigator self-interest has biased estimates of interventions' effectiveness, if not managed properly, this kind of cynicism can lead the public and other scientists to discourage young investigators from even choosing to become experimental criminologists. As articulated below, we need, instead, a culture and set of guidelines that help integrate young investigators' passion to make a difference with the discipline embodied in the conduct of randomized controlled trials.

In order to examine these issues, Eisner notes that we will need a more detailed operationalization of conflict of interest, and more explicit information on the conduct of trials that may lead to bias. The integrity of work in experimental criminology will be improved, and science and policy advanced by heeding Eisner's advice to shine a light on these issues. Given that independent evaluations typically are conducted as effectiveness trials in which interventions are examined under more challenging conditions than efficacy trials, it is important that we simultaneously focus more attention on the implementation process as we examine investigator bias.

Bias or implementation challenges in Eisner's offending studies

Eisner compared the results of trials conducted on four preventive interventions led by developers versus independent evaluators: Reconnecting Youth (RY, a group-based program for youth at high risk of substance use, identified as a model program by the US Substance Abuse and Mental Health Services Administration, SAMHSA); Triple P (a program to promote parenting skills); the Olweus anti-bullying program; and ALERT (a school-based substance-abuse prevention program). Evaluations led by independent evaluators produced much smaller effects than those led by the developers; in the case of RY, independent evaluators found that the program increased the rates of harmful behaviors and that harm increased as fidelity of implementation improved. In reviewing three of the interventions and original trials cited by Eisner (time constraints precluded my review of the Olweus anti-bullying program), I was struck by the following:

First, the original efficacy study of RY was quasi-experimental and flawed fundamentally (Eggert et al. 1994). Substantially higher rates of refusal of individuals to participate in the intervention than those in the control group biased the estimate of program impact. Evidence from a number of studies now suggests that bringing high-risk youth together in groups can have iatrogenic effects (Dishion and Dodge 2005). This problem probably would have been reduced had SAMHSA employed higher evidentiary standards for determining what constitutes a model program.

Second, a large number of randomized controlled trials and quasi-experimental studies have been conducted on Triple P that support its effectiveness when delivered to different populations and in different formats (Nowak and Heinrichs

2008). I reviewed ten randomized controlled trials of Triple P published since 2000 in which the developer was part of the investigative team (Leung et al. 2003; Sanders et al. 2000, 2007; Martin and Sanders 2003; Montgomery et al. 2000; Morawska and Sanders 2006; Plant and Sanders 2007; Roberts et al. 2006; Sanders and Turner 2006; Turner et al. 2007). The results were remarkably positive. What was unclear in these reports was exactly who had executed the randomization, at what stage in the interaction with prospective participants randomization had been conducted, and the degree to which analyses had been conducted with every case randomized. None of the journals in which these findings had been reported required conformity with the Consolidated Standards of Reporting Trials (CONSORT) standards, with one possible exception. It is possible that the differences found between trials conducted by the developer and Eisner have to do with differences in attention to these kinds of fundamental methodological issues.

Third, the original trials of Project ALERT were conducted with randomization at the level of schools; these trials found impacts on children's initiation of cigarette and marijuana use, but not of alcohol, when delivered by teachers (Ellickson and Bell 1990). In the independent replication, Project ALERT was delivered by adults and by adult/teenager teams hired through the state cooperative extension system (St. Pierre et al. 2006). The first report on the independent replication found that the intervention had had no impact on substance use. A subsequent study from the independent evaluators, and not included in Eisner's article, modeled the influence of program leaders' characteristics on program impact (St. Pierre et al. 2007). When groups were led by adults (and to some degree by teenagers) who were more conscientious, sociable, and individuated (confident), Project ALERT had effects on students' cigarette and marijuana use and on their beliefs about alcohol.

This suggests that we need to elevate the standards for conducting and reporting trials and simultaneously to think deeply about how to ensure high-quality program implementation.

Implementation process and intervention effectiveness

Interventions that simply monitor implementation produce effects that are three-times as large as those that do not (DuBois et al. 2002). This says nothing about the influence of actual quality, dosage, or fidelity of implementation, or those underlying factors that contribute to variations in implementation. After reviewing the results of over 500 studies of preventive interventions, Durlak and Dupres (2008) concluded that well-implemented interventions produce effect sizes 2–3 times larger than poorly implemented interventions do; and they generated a list of salient influences on implementation. In order for us to gain a more precise estimate of the role of conflict of interest, we will have to examine simultaneously the multitude of other likely confounding influences that are at least as plausible and important as investigator bias, including:

- Differences in the populations sampled. Variations in population risks and participants' interpretations of the intervention are likely to affect intervention uptake and impact.

- Differences in organizational contexts. Organizations' coordination with other services, positive work environment, openness to change, collaborative decision making, and having a shared vision among staff and leadership are associated with the quality of intervention implementation.
- Differences in community contexts. Historical factors, such as changes in economic conditions (periods of economic growth versus recession) and differences in the availability of other services and policies (such as center-based preschool, welfare reform or three-strike laws) can affect opportunities, community norms, and control-group behavior, which, in turn, can shape the nature of treatment-control contrasts.
- Differences in intervention deliverers' backgrounds. Staff members' perceived needs for, and benefits of, the intervention, training, supports, and proficiency are associated with the quality of intervention implementation.
- Differences in quality of leadership. Organizations and teams led by individuals who are clear in setting priorities, good at facilitating consensus, and skilled in rallying support for an intervention are more successful in implementing interventions than those who are deficient in these respects.
- Differences in support to deliver the intervention well. Training and technical assistance are associated with big differences in the quality of implementation.

The supports needed to ensure that replicated interventions are conducted with quality typically are not well developed. It is not uncommon for investigators to assume that fidelity can be ensured if the intervention is supported in manuals and training procedures, which are necessary, but insufficient. Too little attention is given, for example, to what it takes to produce behavioral change on the part of intervention providers and recipients, particularly to ensure that interventions are relevant and valued (Miller and Rollnick 2002; Rollnick et al. 2007). Similarly, too little attention is given to determining the degree to which a replicated intervention is being developed in an organizational environment capable of supporting it well. These fundamental issues need to be resolved, or at least identified, before useful independent replications can be conducted.

Eisner notes the importance of examining conflict of interest at the same time as implementation, but relevant features of implementation, such as those noted by Durlak and DuPres (2008), are rarely built into the design of intervention replications or examined in depth in effectiveness trials. These factors are likely to affect the degree to which replications of ostensibly the same intervention are actually reproduced across the personal, organizational, and community landscapes in which we conduct our trials. These issues should give pause to intervention developers as they contemplate offering their interventions for public investment; and they should provide guidance to those who conduct meta-analyses and replication trials of interventions. Assessments of conflicts of interest need to be conducted simultaneously with assessments of these other factors.

Independent evaluation

There is ample evidence that some degree of investigator self-interest works its way into science in ways that Eisner articulates so clearly, and, for this reason, independent

replication of intervention findings is a crucial part of the scientific enterprise. A major question for experimental criminologists, however, is when an independent evaluation is appropriate. The types of community-based interventions that many of us develop and examine typically have many moving parts, and these parts need to be aligned properly for interventions to work well. Much of the early replication work that experimentalists do involves refining the intervention, getting the target population identified properly, and learning about organizational and community contextual influences that affect implementation quality (Sherman 2006). Premature independent evaluation of developing interventions that produce negative results may stifle the development of interventions that eventually may be refined and found to work.

Separation of developer and investigator roles?

There are those who behave as if the evidence on developer self-interest is so compelling that the role of program developer ought to be separated from that of evaluator. While Eisner is not proposing this, it is likely that some readers may leap to this conclusion. Petrosino and Soydan (2005) quote Thomas Cook: “Developers are, and should be, passionate advocates for their program, not brokers of honest appraisal.” Until there is more convincing evidence of widespread investigator bias, such a move is imprudent and likely to damage the growing effort to develop, test, and effectively replicate evidence-based interventions. A better approach is to provide young investigators with clear guidance that resonates with their motivation to solve problems and supports their adherence to strong methodological standards.

Some may propose that one way to address this challenge would be for intervention developers to conduct the first trial and for independent investigators to conduct subsequent trials. Our experience is that this is impracticable under current research funding mechanisms and that such a scheme is simply not sufficiently motivating for most individuals drawn to this kind of work. Moreover, it is inefficient for new independent investigators to have to acquire all of the relevant knowledge that the original experimenter learned, while interventions are in their formative stages.

Given the decades-long efforts often required to develop, fund, test, and replicate initial findings, it is crucial that we support the development of intervention scientists who are able to marshal the resources to complete the stages of intervention development, testing, refinement, and replication, with integrity. It is they who possess the energy, commitment, and insight to move such work forward. We need young investigators who possess what Urie Bronfenbrenner (1979) called *disciplined passion*.

I will use our team’s experience to illustrate one way of addressing these issues, and to emphasize why it is so important to have the formative phases of intervention development and replication led by one team. I present it as a way of framing stages of research needed to support the development and replication of effective interventions. In reflecting on our experience, I hope to illustrate some of the challenges that would be created if we were to decompose the experimenter role into developers and evaluators and to create a firewall between them before important questions about intervention design, implementation, impact, and replication are resolved.

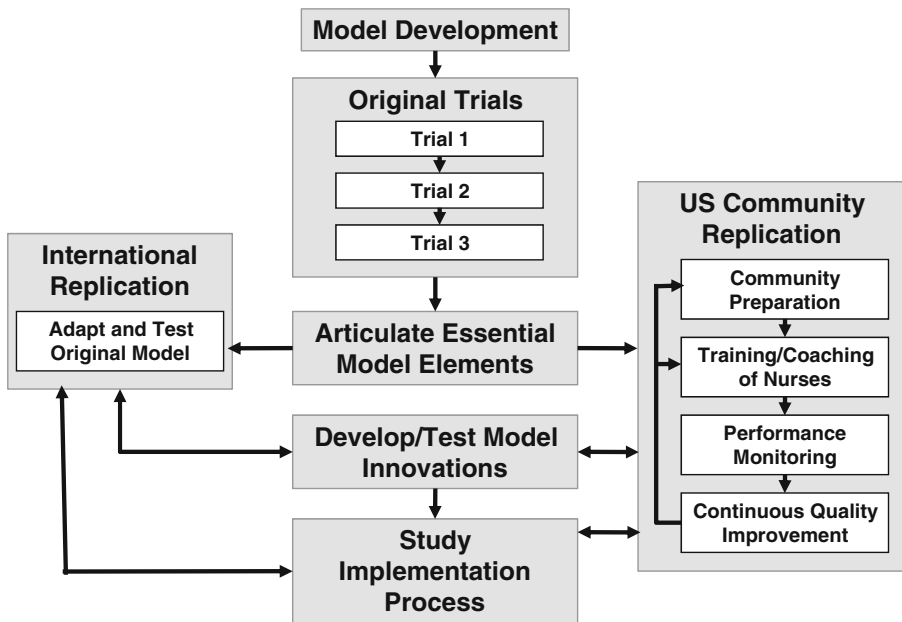


Fig. 1 Model of intervention development, testing, replication, and on-going research and evaluation of the Nurse-Family Partnership

History of trials and expansion of Nurse-Family Partnership

Figure 1 shows a conceptual framework for intervention development and testing that corresponds to the development, testing, and replication of the Nurse-Family Partnership (NFP).

Securing resources to conduct and replicate intervention trials requires money, patience and perseverance We began conducting the original trial of the Nurse-Family Partnership in 1977, in Elmira, New York, USA, focusing on a low-income sample of primarily white pregnant women and their young children, noted in the figure as trial 1. With results of the Elmira trial showing promise, especially for families living in concentrated social disadvantage, in 1984 we decided to determine the extent to which the Elmira findings would be replicated in low-income African-American mothers living in a major urban area. We were concerned about the extent to which the Elmira findings would be replicated with minorities, and some of our findings, such as program impacts on state-verified reports of child abuse and neglect during the child's first 2 years of life, were only trends. We reasoned that the evidence from the first trial was an insufficient foundation for policy and practice.

In 1984 we began to seek funding for the second trial and to identify a community in the USA where such work could be conducted with sufficient numbers of participants and in a health-care delivery system that would allow complete access to mothers' and children's medical records. It took us 4 years and nine funding sources to raise \$7 million to conduct the first phase of replication of the NFP in Memphis, Tennessee (trial 2). The Memphis trial was conducted with a very different population,

with the intervention delivered through a county health department and in the middle of a nursing shortage (half of the nurses quit in the middle of the trial to take higher paying jobs).

Our team would not have invested the energy required to develop this program of research simply to develop and promote a program, because we would not have been satisfied with developing it unless we could assess the degree to which it was working as planned; nor would we have found it gratifying to evaluate others' interventions. What mattered was the excitement of gaining insight into the challenges faced by low-income, socially disadvantaged, mothers and their children and learning whether at least some critical aspects of their lives were changing in accordance with the theoretical models embodied in the program.

Resisting pressure to replicate prematurely Along the way, we resisted an early invitation from the US Surgeon General's office to set up the NFP outside of research contexts—a severely diluted version of the program model that would have allowed the administration to spread around a limited number of dollars to serve a larger number of families. Several years after this invitation, at the end of the 1980s and in the early 1990s, other, less expensive, programs of home visiting for pregnant women and parents of young children were advocated by the National Commission to Prevent Infant Mortality (1989) and the US Advisory Board on Child Abuse and Neglect (1991). These programs, which had no evidence from randomized controlled trials (RCTs) to support their effectiveness, employed paraprofessional home visitors; in fact, early trials of paraprofessional home visiting programs suggested that such models were not working as intended (Olds and Kitzman 1990). In the mid-1990s we began a third trial of the NFP (trial 3) in Denver, Colorado, to examine the relative impact of delivering the program with paraprofessional visitors versus nurses so we might sort out the influence of visitor background on program outcomes (Olds et al. 2002).

Supporting replication with fidelity In 1997 we were invited by the US Justice Department to set up the NFP outside of research contexts in six high-crime neighborhoods. We accepted this invitation because the Memphis trial, in spite of its being conducted with different populations and in substantially different circumstances, was showing evidence of effectiveness on outcomes similar to those examined in the Elmira trial (Kitzman et al. 1997); and a longitudinal study of the Elmira trial was showing long-term effects on state-verified official reports of child maltreatment and on maternal and child arrests (Olds et al. 1998; Olds et al. 1997). Perhaps even more importantly, after 20 years of research, we had begun to gain confidence in our ability to guide nurses in their becoming competent in delivering the essential elements of the program (Refine Essential Model Elements in Fig. 1). Even so, we began the community replication process with apprehension (Olds et al. 2003).

We knew there would be pressures to water down the program and compromise it as we scaled it up in new communities; we did not know whether organizations and communities would commit to delivering the program with fidelity to the model we had tested; and, we did not know whether we could effectively support nurses' delivering the program in new communities. Moreover, we knew from the Elmira

trial that there were sub-populations, such as those experiencing domestic violence (Eckenrode et al. 2000), for whom the program did not work, at least for preventing child maltreatment. Nevertheless, when we compared the results of the Elmira and Memphis trials with other home visiting programs, and when we looked at what was being offered in the way of crime prevention, we reasoned that we had a responsibility to offer the program for limited public investment as long as we simultaneously could build capacity for quality program replication and study the replication process.

We began the community replication process with the belief that we needed to build a system that would help ensure quality program replication. Before an organization can become an NFP site, it must apply and go through an organizational and community assessment process. The organization must sign a contract in which it commits to conducting the program in accordance with program standards, including sending nurses through training, following the program guidelines, and using the program's web-based information system that supports program implementation and serves as the foundation for continuous quality improvement.

Today, the program is being conducted in 350 counties in the USA. A national non-profit-making organization, the Nurse–Family Partnership National Service Office (NFP NSO), has been created to manage the replication of the program (www.nursefamilypartnership.org). As the program is replicated nationally, we have set standards for what evidence can be used as a foundation for public policy as it relates to this program, replicated findings in at least two of the three RCTs. Our research center at the University of Colorado, operating as a kind of research and development unit for the NFP NSO, is now developing program innovations and conducting trials (Trials of Innovations in Model) of those innovations to address vulnerabilities in program design and implementation. These innovations are designed to serve more effectively certain populations (such as families experiencing intimate partner violence), to achieve more effectively certain program objectives (such as involving fathers or promoting competent parental care giving), and to improve fundamentals of implementation (such as retaining families). Throughout all this, the focus of our work is on rigorous improvement of the model and its replication.

Independent evaluation Our team believes that an independent evaluation of the NFP in the USA is important, especially now that the program is being replicated in an increasing number of sites. The question is when. At what point are community replication procedures and model adaptations for different populations developed sufficiently well to give us confidence that testing the program would give a fair test? Our team is conducting research to make sure that the replication system and model are operating as well as possible, while others decide when an independent replication is warranted.

International replication In recent years we have been invited to set up the NFP in societies outside of the USA. As the NFP has been replicated internationally, we have established a set of principles to guide this work: First, we make no claims about the impact of the program in other societies, where populations and health and human service delivery systems may differ so much that we cannot generalize from

the US experience. Second, we recommend a set of stages through which international replication must go in order to warrant large scale public investment in the program: (a) pilot work to adapt the NFP to local circumstances and to determine its acceptability and feasibility; (b) evaluation to determine its likelihood of producing effects of public health importance; (c) a well-powered RCT conducted by an independent team of investigators from the local society to determine if impacts on functional and economic outcomes justify investments in the program on a larger scale; and (d) a commitment to conduct the program with fidelity to the model tested if it is found to work in the trial. The challenge for us is how to support international program replication in an evidence-based, responsible, way.

Thoughts for new experimental criminologists

How do we support the development of experimental criminologists in ways that enable them to stay true to their commitment to solve problems in rigorous ways? Here are some specific thoughts for new experimental criminologists about how to navigate these challenging roles:

1. Clarify your focus. Are you in this business to promote an intervention or to solve a problem? If you are entering this field to prove that your intervention works, you are vulnerable to the conflict of interest that Eisner identifies. If you enter this field to solve a problem, you have an internal compass that will help guide your decision making and reduce your susceptibility. Keep your focus.
2. Take the time to develop your intervention before you test it and before you replicate it. A substantial portion of the failure of work in intervention research can be attributed to premature testing and replication. Respect the challenges of producing and sustaining changes in behavior and context by devoting the time to develop an intervention that is robust.
3. Expect that your first trial may be as much a clinical inquiry as a focused clinical trial (Tukey 1977). The extent to which you can formulate strong hypotheses in your trials will depend upon how much prior work has been conducted with the intervention you are testing. Expect and embrace heterogeneity in response, but remain skeptical about patterns of moderated intervention effects, especially if you did not predict them, if subgroups are small, or if there is no coherent explanation for the pattern of results. Your first trial will serve as the foundation for formulating subsequent hypotheses.
4. Focus your analysis on those outcomes specified in your original hypotheses; operationalize the outcomes before you begin the analysis and make sure that all primary outcomes are reported. If you have not followed these rules, say so, as the statistical significance of your findings will be affected.
5. Commit to conducting a replication trial to discern the extent to which intervention main effects and moderated effects are reproducible. Given the cost and length of time it takes to conduct intervention research in community settings, it is highly unlikely that you will conduct replications in exactly the same circumstances as your original work, as one might in laboratory research.

You are likely to want to determine the extent to which your intervention is equally effective when delivered in different contexts. Does it work with populations that vary from the one you examined in your first trial? How about in different service delivery contexts? These are fundamental questions regarding intervention generalizability, ones that will be difficult to address as thoroughly as any of us would like, given that so many factors will change simultaneously with each new replication. If there is consistency in impact, it suggests that the intervention is robust. If not, it is important to understand why, as this will provide leads about how to strengthen the intervention.

6. Expect that replication trials will be difficult to interpret. When you change the study parameters and find that the intervention is not working, you will not know whether failure in subsequent trials is due to changes in population, organizational context, historical factors, or differences in adequacy of data derived from different administrative systems, etc. Nevertheless, replication is needed, as is careful attention to those factors that may explain variation in response as the intervention is transported to a new trial. Without replication of key outcomes, the intervention is not ready for public investment.
7. Remove your personal financial interests from the success of the research. Rigorous research that produces promising results is likely to lead to additional funding and personal acknowledgment; this is both unavoidable and desirable. Do not compromise the interpretation of your results, however, by accepting personal financial gain from marketing your program. It is not consistent with solving the problem; it may cloud your judgment; and it ultimately will compromise your achieving your goal.
8. Understand the challenges of conducting trials in community settings and adhere to the highest standards for their conduct. Research standards are improving all the time. The durability of your work will depend upon its integrity. One way of addressing this issue is to make biostatisticians your colleagues, from the earliest stages of planning, conducting, and analyzing your trials.
9. Establish outcomes that are of public health/clinical/social importance. Policy makers and practitioners will be interested in outcomes that have relatively unambiguous public health meaning and that relate to their policy agendas.
10. Resist invitations to offer your intervention for public investment before it is ready. Given the paucity of good research to guide policy and practice, there will be lots of pressure to replicate your intervention before the science and practice have been sufficiently well developed. Caving into that pressure will undermine your credibility and, possibly, lead to the expansion of an intervention that does not work. It is bad for the population you wish to help, for tax payers, and for your career.

Thoughts for editors

Perry and Johnson (2008) have recommended that the CONSORT statement published originally in 1996 for medical journals be applied to criminal justice journals. I agree. This should increase the transparency of reports of trials, which

may help reduce the gap between findings from original trials and replication studies. It is also important for journals to identify stables of reviewers who understand the conduct of field experiments. We all have seen reports of trials that violated fundamental standards in research design and implementation, which could have been addressed with better review.

Thoughts for those who conduct systematic reviews

In examining the impact of interventions across trials, it is important to keep in mind that heterogeneity in populations served, community and organizational contexts, history, and quality of implementation are likely to moderate intervention impact. Standard methods of examining heterogeneity in meta-analyses are not likely to detect such moderation, especially when intervention impact depends upon the presence of several conditions at once (e.g., correct specification of target population, a carefully delivered program model, intervention delivered by providers with necessary background). Moreover, these features of trials may be confounded with one another (and with the developer/independent evaluator role), making the interpretation of systematic reviews difficult.

Thoughts for mentors/training programs

Given the challenges of conducting randomized trials in community settings, it is crucial that prevention scientists have formal training in the conduct of RCTs and that they be given additional training in conducting trials in complex community settings. This will require linkages with departments of biostatistics in schools of medicine or public health, where the conducting of trials is much more common. Additional attention should be focused on conflicts of interest and how these conflicts can affect the design, conduct, and interpretation of trials.

Policy makers

One of the challenges facing scientists who conduct intervention research is that expectations have been raised to unrealistic heights for some policy makers and parts of the public. We should help policy makers become better consumers of evidence, develop realistic expectations for what public investments in evidence-based interventions are likely to achieve, and understand why investments in research and replications of evidence-based practices hold the greatest potential for addressing social ills.

Conclusions

Manuel Eisner has performed a useful service for experimental criminology by posing the cynic's hypothesis, synthesizing evidence to show why self-interest is a significant concern, and proposing a series of research strategies to address this

issue. While he acknowledges the need to study conflict of interest and implementation quality at the same time, the assessment of implementation quality is not as straightforward nor the depth of information in existing databases as extensive as he suggests. A strong case can be made for deepening our measurement and understanding of implementation and factors that influence it, while examining developer-as-evaluator conflict of interest. Understanding these factors is important for sorting out the influence of conflict of interest and for guiding successful policy and practice. Given the likelihood that conflict of interest compromises estimates of treatment impact to some degree, independent evaluation is crucial, but premature independent evaluation may stifle interventions before they have reached maturity and warrant independent evaluation. There have been calls for the separation of the role of developer and evaluator, which, if enacted, would impede the development of effective interventions. A better strategy is to follow Eisner's recommendations for studying conflict of interest, to raise the bar for conducting and reporting trials, and to provide young investigators with a roadmap for how to integrate their passion to make a difference with the discipline embodied in the conduct of randomized controlled trials.

Acknowledgments Dennis Luckey, Mike Knudtson, Nancy Donelan-McCall, Harriet Kitzman, and Charles Henderson provided helpful comments.

References

- Bronfenbrenner, U. (1979). *The ecology of human development: experiments by nature and design*. Cambridge, MA: Harvard University Press.
- Dishion, T., & Dodge, K. (2005). Peer contagion in interventions for children and adolescents: moving towards and understanding of the ecology and dynamics of change. *Journal of Abnormal Child Psychology*, *33*(3), 395–400.
- DuBois, D. L., Holloway, B. E., Valentine, J. C., & Cooper, H. (2002). Effectiveness of mentoring programs for youth: a meta-analytic review. *American Journal of Community Psychology*, *30*(2), 157–197.
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*(3–4), 327–350.
- Eckenrode, J., Ganzel, B., Henderson, C. R., Jr, Smith, E., Olds, D. L., Powers, J., et al. (2000). Preventing child abuse and neglect with a program of nurse home visitation: the limiting effects of domestic violence. *JAMA*, *284*(11), 1385–1391.
- Eggert, L. L., Thompson, E. A., Herting, J. R., Nicholas, L. J., & Dickers, B. G. (1994). Preventing adolescent drug abuse and high school dropout through an intensive social network development program. *American Journal of Health Promotion*, *8*(2), 202–215.
- Eisner, M. (2009). No effects in independent prevention trials: can we reject the cynical view? *Experimental Criminology*.
- Ellickson, P. L., & Bell, R. M. (1990). Drug prevention in junior high: a multi-site longitudinal test. *Science*, *247*, 1299–1305.
- Kitzman, H., Olds, D. L., Henderson, C. R., Jr, Hanks, C., Cole, R., Tatelbaum, R., et al. (1997). Effect of prenatal and infancy home visitation by nurses on pregnancy outcomes, childhood injuries, and repeated childbearing. A randomized controlled trial. *JAMA*, *278*(8), 644–652.
- Leung, C., Sanders, M. R., Leung, S., Mak, R., & Lau, J. (2003). An outcome evaluation of the implementation of the Triple P-Positive Parenting Program in Hong Kong. *Family Process*, *42*(4), 531–544.

- Martin, A. J., & Sanders, M. R. (2003). Balancing work and family: a controlled evaluation of the Triple P-positive parenting program as a work-site intervention. *Child and Adolescent Mental Health, 8*(4), 161–169.
- Miller, W. R., & Rollnick, S. (2002). *Motivational interviewing, second edition: preparing people for change*. New York: Guilford Press.
- Montgomery, D. T., Sanders, M. R., & Brechman-Toussaint, M. L. (2000). The mass media and the prevention of child behavior problems: the evaluation of a television series to promote positive outcomes for parents and their children. *Journal of Child Psychology and Psychiatry, 41*(7), 939–948.
- Morawska, A., & Sanders, M. R. (2006). Self-administered behavioral family intervention for parent of toddlers: Part I. Efficacy. *Journal of Consulting and Clinical Psychology, 74*(1), 10–19.
- National Commission to Prevent Infant Mortality. (1989). *Home visiting: opening doors for America's pregnant women and children*. Washington, D.C.: National Commission to Prevent Infant Mortality.
- Nowak, C., & Heinrichs, N. (2008). A comprehensive meta-analysis of Triple-P parenting program using hierarchical linear modeling: effectiveness and moderating variables. *Clinical Child and Family Psychology Review, 11*, 114–144.
- Olds, D. L., & Kitzman, H. (1990). Can home visitation improve the health of women and children at environmental risk? *Pediatrics, 86*(1), 108–116.
- Olds, D. L., Eckenrode, J., Henderson, C. R., Jr, Kitzman, H., Powers, J., Cole, R., et al. (1997). Long-term effects of home visitation on maternal life course and child abuse and neglect. Fifteen-year follow-up of a randomized trial. *JAMA, 278*(8), 637–43.
- Olds, D., Henderson, C. R., Jr, 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. *JAMA, 280*(14), 1238–1244.
- Olds, D. L., Robinson, J., O'Brien, R., Luckey, D. W., Pettitt, L. M., Henderson, C. R., Jr, et al. (2002). Home visiting by paraprofessionals and by nurses: a randomized, controlled trial. *Pediatrics, 110*(3), 486–496.
- Olds, D. L., Hill, P. L., O'Brien, R., Racine, D., & Moritz, P. (2003). Taking preventive intervention to scale: the Nurse-Family Partnership. *Cognitive and Behavioral Practice, 10*(4), 278–290.
- Perry, A. E., & Johnson, M. (2008). Applying CONSORT to criminal justice trials: associations between journal impact factors and quality of report. *Journal of Experimental Criminology, 4*(2), 168–185.
- Petrosino, A., & Soydan, H. (2005). The impact of program developers as evaluators on criminal recidivism: results from meta-analyses of experimental and quasi-experimental research. *Journal of Experimental Criminology, 1*(4), 435–450.
- Plant, K. M., & Sanders, M. R. (2007). Reducing problem behavior during care-giving in families of preschool-aged children with developmental disabilities. *Research in Developmental Disabilities, 28*, 362–385.
- Roberts, C., Mazzucchelli, T., Studman, L., & Sanders, M. R. (2006). Behavioral family intervention for children with developmental disabilities and behavioral problems. *Journal of Clinical Child & Adolescent Psychology, 35*(2), 180–193.
- Rollnick, S., Miller, W. R., & Butler, C. C. (2007). *Motivational interviewing in health care: helping patients change behavior*. New York: Guilford Press.
- Sanders, M. R., & Turner, K. M. T. (2006). Help when it's needed first: a controlled evaluation of brief, preventive behavioral family intervention in a primary care setting. *Behavior Therapy, 37*, 131–142.
- Sanders, M. R., Markie-Dadds, C., Tully, L. A., & Bor, W. (2000). The Triple P-Positive Parenting Program: a comparison of enhanced, standard, and self-directed behavioral family intervention for parents of children with early onset conduct problems. *Journal of Consulting and Clinical Psychology, 68*(4), 624–640.
- Sanders, M. R., Bor, W., & Morawska, A. (2007). Maintenance of treatment gains: a comparison of enhanced, standard, and self-directed Triple P-positive parenting program. *J Abnormal Child Psychol, 35*, 983–998.
- Sherman, L. W. (2006). “To develop and test:” the inventive difference between evaluation and experimentation. *Journal of Experimental Criminology, 2*(3), 393–406.
- St Pierre, T. L., Osgood, D. W., Mincemoyer, C. C., Kaltreider, D. L. & Kauh, T. J. (2006). Results of an independent evaluation of project alert delivered in schools by cooperative extension. *Prevention Science, 6*(4), 305–317.
- St Pierre, T. L., Osgood, D. W., Siennick, S., Kauh, T., & Burden, F. (2007). Project ALERT with outside leaders: what leader characteristics are important for success? *Prevention Science, 8*(1), 51–64.
- Tukey, J. W. (1977). Some thoughts on clinical trials, especially problems of multiplicity. *Science, 19*, 679–684.

- Turner, K. M. T., Richards, M., & Sanders, M. R. (2007). Randomised clinical trial of a group parent education programme for Australian indigenous families. *Journal of Pediatrics and Child Health, 43*, 243–251.
- US Advisory Board on Child Abuse and Neglect. (1991). *Creating caring communities: blueprint for an effective federal policy on child abuse and neglect*. Washington, DC: US Government Printing Office.

David Olds is professor of pediatrics, psychiatry, and preventive medicine at the University of Colorado where he directs the Prevention Research Center for Family and Child Health.