

# Commentary on Valentine, Jeffrey, et al. *Replication in Prevention Science*

## The Advisory Board of Blueprints for Violence Prevention

Steve Aos • Thomas D. Cook • Delbert S. Elliott •  
Denise C. Gottfredson • J. David Hawkins •  
Mark W. Lipsey • Patrick Tolan

Published online: 15 May 2011  
© Society for Prevention Research 2011

The Valentine et al. paper provides a first-step analysis of some key issues in replication. The paper acknowledges the question of what is to be considered a replication and introduces the importance of understanding the purpose of a replication study. Importantly, the paper recognizes the current development of the field of prevention science in which we are faced with few replications to consider, with sometimes a maximum of two studies. To illustrate, the

Blueprints for Violence Prevention project at the University of Colorado has reviewed evaluations of 895 prevention programs. Of these, less than 13% have been subjected to a second study, only some of which would likely have been considered true replications. The Valentine et al. paper focuses primarily on how information should be summarized across multiple studies, where inclusion as replication is not uncertain, and where the needed information to judge replication by the particular standard of interest is accessible. We agree with the paper's conclusions about how evidence from multiple studies that meet these criteria should be summarized to answer the question, "What does the available evidence say about the size of the effect attributable to Intervention A?"

While the methods proposed in the paper are a useful point of departure for answering the question posed, the paper largely ignores an important question about the type of information needed by policy makers to make decisions about program effectiveness given the current stage of development of prevention science. We agree with the paper's conclusion that "inferences about replication can be made with as few as two studies, but only within a very weak inferential framework". However, we also recognize, as do the paper's authors, that it is not feasible to wait until we have enough information to make strong inferences before recommending adoption of prevention programs to policy makers who are seeking to fund and advance the use of evidence-based prevention programs rather than programs without any scientific evidence of effectiveness. Given the current state of prevention science, efforts that attempt to identify effective programs or practices for wider dissemination (such as Blueprints for Violence Prevention) seldom rely on estimates of the exact size of the effect. Instead, they attempt to identify programs that have a non-

---

S. Aos  
Washington State Institute for Public Policy,  
Olympia, WA, USA

T. D. Cook  
Institute for Policy Research, Northwestern University,  
Evanston, IL, USA

D. S. Elliott  
Center for the Study and Prevention of Violence,  
Institute of Behavioral Science, University of Colorado,  
Boulder, CO, USA

D. C. Gottfredson  
Department of Criminology and Criminal Justice,  
University of Maryland,  
College Park, MD, USA

J. D. Hawkins (✉)  
Social Development Research Group, School of Social Work,  
University of Washington,  
Seattle, WA, USA  
e-mail: jdh@u.washington.edu

M. W. Lipsey  
Peabody Research Institute, Vanderbilt University,  
Nashville, TN, USA

P. Tolan  
Center for Positive Youth Development, University of Virginia,  
Charlottesville, VA, USA

zero effect that is reliable. This issue about the nature of the evidence needed by policy makers to make a decision given the current state of development of prevention science is not adequately addressed in the Valentine et al. paper. Currently, what is needed is a statement of the minimum requirements of replication studies to allow a tentative conclusion about the durability or reliability of non-zero preventive effects.

The paper misses the opportunity to operationally define “replication” and to set standards for how replication studies should be designed for the field. The issue of what the replication is meant to demonstrate is most basic. Is the replication intended to provide assurance that a specific program, if offered as designed to the population(s) with which it has been tested, can be expected to reliably have a non-zero effect on the desired outcome in future applications? Is it to determine if a class of programs (e.g., mentoring for populations at risk for delinquency) has some significant benefits? Is it to determine from a large class of interventions the factors that make a difference (as illustrated by the Lipsey (2009) and Weisz et al. (2006) meta-analyses of interventions with adolescents)? Or is replication undertaken to demonstrate that a guiding theory is not rejected/is supported (e.g., programs that focus on problem-solving skills and increase self-control lead to lower rates of problem behavior)? In our view, the paper would have been more helpful if it had elaborated on the types of replication research described and developed standards for the design of each type of replication study. At this point in the development of prevention science, it would have been especially helpful for the paper to provide standards for replication studies that seek to answer the first question above, that is “Can a specific program, if offered as designed, be expected to reliably produce a non-zero effect on the desired outcome in future applications with the population(s) with which it has been tested?”

It also would have been helpful if the paper had discussed more completely the different dimensions that are likely to vary in the real world in replication studies (e.g., location, population characteristics, developer involvement, research team decisions, implementation quality) and that might influence program effectiveness. This discussion could then have led to standards for the design of future replication studies. Though touched in the section

“Can one study be considered a replicate of another?”, we would like to have seen more discussion of the thorny issue of how to judge whether study B is a replication of study A in light of differences in the content of programs from one study to another. That is, programs are often modified before being subjected to new evaluation. The field needs guidelines to help determine how much and what types of changes to a program should cause the new evaluation to be considered a study of a different intervention rather than a replication of the original. Further, the paper did not explore another important possible difference in replication studies: outcome measurement. What should be done when two or more studies use two or more similar—but not exactly the same—outcomes (e.g., arrests, convictions, self-reports of crime)? What if different outcome measures used in different studies produce conflicting estimates of outcomes with one measure suggesting an effect but other measures suggesting no effect? How should this conflicting information from different outcome measures be weighed in making a decision about whether or not an effect is replicated?

In sum, while this paper makes an important contribution in suggesting methods for combining results from multiple studies to determine the effect size likely to be achieved by an intervention, we were disappointed that it did not achieve the larger objective of providing standards to guide replications in prevention science in the future. Prevention scientists still need to begin discussing specific “road rules” to be followed to guide a decision about whether to include a study as a replication study or not and about how to combine studies that are considered replications. We compliment the authors on initiating a conversation on the important topic of replication, and we look forward to discussions that will create a viable operational definition of replication for the field of prevention science.

## References

- Lipsey, M.W. (2009). The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims and Offenders, 4*, 124–147.
- Weisz, J.R., McCarty, C.A., & Valeri, S.M. (2006). Effects of psychotherapy for depression in children and adolescents: A meta-analysis. *Psychological Bulletin, 132*, 132–149.