Commentaries on *Replication in Prevention Science*: A Rejoinder

Jeffrey C. Valentine • Anthony Biglan • Robert F. Boruch • Felipe González Castro •
Linda M. Collins • Brian R. Flay • Sheppard Kellam • Eve K. Mościcki •
Steven P. Schinke

Published online: 7 May 2011
© Society for Prevention Research 2011

“People have turned to their senses for information about the world on which to ground their natural philosophy, their medicine and such, so long as there have been such disciplines. And so long as there have been such disciplines, halfway reflective people must have worried to at least some degree about ... how we can be sure that the individual and particular observations we make on a given occasion are not misleading in some way...To this apparently simple question, we find in much early literature a relatively simple answer...Do it again.” (Garber 1997, pp. 25–26).

Our paper “Replication in Prevention Science” (Valentine et al. 2011) was designed, in part, to provoke conversation about the matter of replication, its meanings, its desirability (within limits), the limitations in our understandings of it, and how to get it (under our definitions) when it is sensible

to try. We are grateful to the Editor for inviting commentary and also for inviting a rejoinder. The reactions in the commentaries are plainly put. Further, they are constructed in the same scientific spirit that drove our own questions about the possibility, partial possibility, or impossibility of replication, and have articulated some of the continuing challenges faced by prevention scientists in the development, testing, and dissemination of effective prevention strategies. We are pleased that the paper provoked conversation so briskly, and among very able colleagues.

In their commentary on “Replication in Prevention Science,” Aos et al. (2011) are courteous and thoughtful, for instance, in declaring that it would have been helpful if we had done things a bit differently. Despite this complaint, several sources of agreement between their commentary and our paper are evident. For example, we

J. C. Valentine (✉)
University of Louisville,
Louisville, KY, USA
e-mail: jeff.valentine@louisville.edu

A. Biglan
Oregon Research Institute,
Eugene, OR, USA

R. F. Boruch
University of Pennsylvania,
Philadelphia, PA, USA

F. G. Castro
Arizona State University,
Phoenix, AZ, USA

L. M. Collins
Pennsylvania State University,
University Park, PA, USA

B. R. Flay
Oregon State University,
Corvallis, OR, USA

S. Kellam
American Institutes for Research,
Washington, DC, USA

E. K. Mościcki
American Psychiatric Institute for Research and Education,
Arlington, VA, USA

S. P. Schinke
Columbia University,
New York, NY, USA

agree that policy makers are often not in a position to wait until an evidence base supports strong inferences to make decisions. At the same time, there is ample reason to believe that people in general use inferential strategies to make sense out of a body of evidence in ways that are likely to confuse rather than to clarify issues. Consequently, one of the major recommendations in Valentine et al. (2011) is that state-of-the-art literature synthesis techniques ought to be applied to all synthesis questions, even if very few studies are thought to exist that pertain to the research question.

The question that Aos et al. (2011) would liked us to have addressed can be framed as “How [does one] judge whether study B is a replication of study A given differences in content of programs from one study to another?”. They acknowledge that this is a thorny issue, and wish that we would have expanded on our treatment of it. Indeed, we started to address the matter in the section headed “*Can One Study Be Considered a Replicate of Another?*”. Hansen (2011) expresses pessimism that posing this question even makes sense. We will try again to pose the question in other ways.

Imagine the following problem (Glass 2000, adapted from Nozick 1981, who in turn adapted it from Plutarch 75/1957): Theseus roams the Mediterranean Sea in a ship. Every day, one of the ship’s planks is removed, taken ashore, and replaced with a new plank. The old planks are then used to assemble an exact replica of the original ship. After enough time, every plank that Theseus originally left port with has been replaced, and all of the original planks are part of a vessel that is now sitting in dry dock. The interesting question is whether the ship on which Theseus is currently sailing is “the same” as the ship with which he originally set to sea.

The point is this: Strictly speaking *nothing* is ever “the same” (in Herodotus’s terms, both the river and the person are continually changing). Therefore, in order to make any sense at all the question about identity has to be reframed to ask whether two different things are qualitatively “similar enough” to be labeled “the same,” a question for which no *a priori* answer exists. For Glass (2000) in particular, the answer depends on how much weight is assigned to the different components of the object. For example, assume a preventive intervention has a mentoring component in study A but not in study B. If the mentoring component is presumed to be important then we would likely say that B is not a replication of A. If however the mentoring component is thought to be peripheral, we may be more inclined to call B a replication of A. All of this implies that the person making the judgment of similarity must possess a deep understanding of the theoretical and contextual issues surrounding a preventive intervention before there is much hope of making reasonable judgments regarding the

very difficult question of whether one study can be considered to be a replication of another.

Colleagues Aos et al. (2011) continue their critique by pointing out the many interesting (and potentially frustrating) ways that studies might differ from one another. For example, they note that the outcome can be operationalized in different ways across studies, and that this can make inferences more difficult. We concur. Indeed, it is for this reason that Valentine et al. (2011) called for replication research that is more intentional and, moreover, discussed structures that could be set up to support that kind of replication research. We take the opportunity here to point out that these judgments will not always be difficult. For example, if one study of a smoking cessation intervention uses self-reports as the outcome and the other uses chemical analysis of saliva, then all else being equal, interpretative priority ought to be given to the study that used the better outcome measure. What is “better” of course is a matter of science and research policy.

Despite our assertion that the question posed by Aos et al. (2011) is extraordinarily difficult to address in the abstract, we agree with these commentators that policy makers do need something concrete to go on in making decisions about preventive interventions. We do not claim to have ready answers, and believe that how best to do this should be near the top of the list in the conversation provoked by our paper (as a start, we offer the section “Statistical Options for Assessing the Results of a Small Number of Studies” and Table 1 from the “Replication in Prevention Science” paper). And, as we suggested in the paper, one potentially helpful way of framing the question of whether one study can be considered to be a replicate of another is to consider the theory of action underlying the preventive intervention. Further, qualitative studies that explore how experts make judgments of similarity ought to be helpful.

Commentator Hansen (2011) drills down more deeply into the dimensions of failure to replicate and to the boundary conditions that delimit the possibility of replication. We appreciate Hansen’s comments too. In particular, it is worth repeating his assertion that “replication by itself will not ensure the advancement of the field”. We further agree with Hansen (2011) that implementation fidelity and adaptation speak to core issues in any kind of intervention research and its replication. For example, with respect to adaptation the question of how much one may modify an intervention and still expect to get a reasonably similar effect is vitally important, and at the same time in no way limited to prevention science. Of course, another way to frame this question is to investigate the core elements of an intervention that cannot be changed without altering its fundamental nature. To illustrate, consider the pharmaceutical industry. When a patent expires on a specific

medication a "new generation" medication is often developed by changing, say, one molecular detail. The new one is then marketed under a new name, with the old version being sold as a generic and the new as a proprietary drug. This is clearly another case where the weighting of what matters is important, and that weighing depends on how the problem is framed. From the perspective of the drug company, it is important to say that the medications are "different." From the standpoint of a person with limited resources who needs the medication, the small change might not matter nearly so much as the presumed efficacy of both drugs and their relative costs.

These substantial agreements aside, we believe that Hansen (2011) is unduly pessimistic about the value of replication. This pessimism manifests itself in two ways. First, in an argument that we partially addressed above, he asserts that secular trends mean that interventions must continually adapt or risk being out-of-date. It follows that if interventions are constantly changing it is difficult to replicate them. Our agreement with this concern should not be read as agreement with the implication that we can not or do not need to attempt to replicate research. Furthermore, Hansen's conclusion seems to be tied to a particular level of analysis that focuses on the specific tactics that might be used to address a problem. As opposed to specific tactics, we believe that general strategies are less likely to be prone to secular trends. For example, it is well known that one good way to change a person's attitude is to get them to try to change someone else's attitude. It is not sufficient to coerce the person to try to change someone else's attitude: The trick is to get the person whose attitude you want to change to volunteer to try to change the other person's attitude. The specific ways that this might be accomplished are likely to vary quite a bit as a function of circumstance, but the general strategy (that trying to change someone else's attitude must be perceived as a volitional act) is less likely than specific tactics to require modification.

Second, one interpretation of Hansen's (2011) critique is that replication will not have value until prevention science solves the problems of implementation fidelity and how to adapt interventions for use in other contexts. However, it seems clear that if this is true about replication research then it is equally true about the original experiments as well. Practically no one is interested only in the narrow question of whether a specific implementation of an intervention had an effect in a specific context on a specific group of people. Instead, we virtually always hope for

insight into the likelihood that results observed in one study will generalize to variations in study conditions. There is little value in experimentation without replication: They go hand in glove.

Finally, we note that Hansen (2011) fails to provide an alternative to replication. Are policy makers better off using Campbell's (1986) rule of proximal similarity, with all of the ambiguous judgments that doing so entails? Or should they simply do what they think is best? As we argued in the paper, we think that relying on state-of-the-art synthesis techniques is generally more likely to lead to better conclusions regarding the "likely direction, size, and plausible range of intervention effects as suggested by the current knowledge base". In this regard, we strongly agree with Hansen's concluding remark that "the goal of understanding how and why we succeed and fail ... should be enshrined".

We want to close this note by thanking, once again, the individuals who contributed their time and energy to providing thoughtful commentary on our paper. We also want to circle back in a way that keeps our eyes on the prize: We are all interested in moving forward prevention science not merely for its own sake, but for the sake of improving public health. Being at least "halfway reflective people" we are, as Garber suggested, worried about whether "the individual and particular observations we make on a given occasion are not misleading in some way." Our suggestion is to experiment and "do it again," and then to use the best synthesis techniques possible to try to make sense of the world.

References

- Aos, et al. (2011). Need exact title and author list. *Prevention Science*. doi:10.1007/s11121-011-0219-4
- Garber, D. (1997). Experiment, community, and the constitution of nature in the seventeenth century. In J. Earman & J. D. Norton (Eds.), *The cosmos of science: Essays of exploration* (pp. 24–54). Pittsburgh, PA: University of Pittsburgh Press.
- Glass, G. V. (2000). *Meta-analysis at 25*. Retrieved from <http://glass.ed.asu.edu/gene/papers/meta25.html>.
- Hansen, E. B. (2011). Was Herodotus correct? *Prevention Science*. doi:10.1007/s11121-011-0218-5
- Nozick, R. (1981). *Philosophical explanations*. Cambridge, MA: Harvard University Press.
- Plutarch (75/1957). *The lives of the noble Grecians and Romans*. (J. Dryden, Translator). New York: The Modern Library.
- Valentine, J. C., Biglan, A., Boruch, R. F., Castro, F. G., Collins, L. M., Flay, B. R., et al. (2011). Replication in prevention science. *Prevention Science*. doi:10.1007/s11121-011-0217-6