Published in final edited form as:

Subst Use Misuse. 2012 ; 47(0): 1436–1438. doi:10.3109/10826084.2012.719723.

Robert Hornik and Robert Orwin on Stephen Magura's “Failure of Intervention or Failure of Evaluation: A Meta-Evaluation of the National Youth Anti-Drug Media Campaign Evaluation”

Robert Hornik, Ph.D [Wilbur Schramm Professor of Communication and Health Policy] and
Annenberg School for Communication, University of Pennsylvania

Robert Orwin, Ph.D [Senior Study Director]
Substance Abuse Study Group, Westat

1. The basic design concern

Dr. Magura believes that our use of a longitudinal panel observational design was problematic and perhaps unethical. We should have been doing some form of cluster randomized design, instead. His argument is that our results were guaranteed to be useless, because we could never be sure that all confounders (measured and unmeasured) were balanced between more and less exposed youth. Our counter-argument is as follows: (a) we had no choice – none of the relevant political actors had any interest in spending \$1 billion on an experiment – so we had to decide whether we could learn something valuable from a non-experimental design. There were specific discussions about the need for such a cluster randomized design before the RFP for the evaluation was developed, and they were specifically rejected by ONDCP and NIDA. As we note below, we think that was a correct decision. (b) He argues that the government should not have been investing in the campaign without adequate a priori experimental evidence of its effects. This argument ignores the actual context of the decision: while the government involvement in the campaign was new, the actual campaign had been operating for 20 years under the sponsorship of the Partnership for a Drug Free America (PDFA) – whose ability to obtain free advertising time had been reduced – and which had gone to Congress for support on the basis of its own long term trend data and its political clout. There was some experimental evidence for success for a related campaign (Palmgreen et al.) In that sense this was an evaluation of an operational program already assumed to be effective, not of a novel intervention about which there was doubt in the policy community. The decision to fund an independent evaluation to check up on the value of the operational program was the notable decision. (c) It is not obvious that a cluster randomized design would have been helpful, here. Such designs look ideal on paper and for some forms of interventions they may be useful in practice; for large scale communication programs, success may not depend merely on individual response to persuasive messages but on social network diffusion of messages and responses of institutions over the long term and on a large scale. Such designs, as typically implemented,

Each of Dr. Magura's specific critiques is right in the abstract, and in our view deeply wrong in the practice of evaluation. Our comments here will be both specific - direct responses to the elements of his critique, and general – a statement of contrasting philosophies of evaluation.

may provide a very good answer to the wrong question. This idea is developed elsewhere (Hornik, 2002). (d) As we argue in the paper and below, the design we did implement was a reasonable alternative; it was informative, in the sense that its results would provide some guidance to policy decisions, compared to no evaluation, or a less careful or biased evaluation. [While the cluster randomized design would have been a superficially attractive but in practice a weaker alternative, a stronger alternative might have been measuring advertising time purchased across metropolitan areas (rather than personal claims of exposure to ads) and using that as an exogenous, although aggregate, estimate of exposure, examining whether that was associated with declines in drug use at the community level; unfortunately the actual variation in exposure across communities was too small to implement this form of analysis.

2. Our measurement of drug use

Dr. Magura is skeptical that we captured actual drug use with our self-report measures, which were likely subject to social desirability. He would have preferred that we use physiological measures – urine tests. The counterarguments: (a) This is the standard way the government has been measuring youth drug use for a very long time, including the central policy foundations: the Monitoring the Future Study, and the National Survey on Drug Use and Health. The decision to launch this campaign was based on concerns deriving from those surveys. Take them on if you want, but don't be unrealistic about what is the standard now. (b) We did our best to establish that our measures of outcomes were valid – in particular showing the high level of consistency in self-reports of use over multiple year-apart waves of data collection, the high level of predictive power of intentions to use marijuana at one wave on reports of initiation of use in a subsequent wave, and the consistency in a wide range of beliefs and attitudes, norms, efficacy to refuse, intentions and behavior, at any given wave. (c) We had no other reasonable choice – we never thought we would be able to obtain agreement from a nationally representative sample of youth and retain them in a cohort study for four years if we were collecting urine from them every time we interviewed them. (d) Given the results of the evaluation – and no evidence that exposure to the campaign was positively associated with cognitions or behavior, in practice social desirability is not a credible competing hypothesis for the observed results.

3. The analysis approach

Dr. Magura argues that our results would have been uninterpretable had we found positive effects of the campaign. By that we understand that he means that we could not have made a definitive claim of positive effects and thus we might as well have not bothered. As he notes, a priori we set three criteria for success: positive secular trend, cross-sectional association between exposure and outcomes, adjusted for measured confounders, and lagged association between prior exposure and subsequent outcomes adjusted for prior outcomes and measured confounders. Despite the framing of the critique which suggested that we seemed to be ignorant of the threats to inference associated with each of these analyses, of course the logic of why we wanted to use all three approaches was precisely the one he outlines and that we were explicit about: each approach faced different threats (history and maturation in the secular trend analysis; unmeasured confounders and causal direction for

the cross-sectional analyses; unmeasured confounders for the lagged analyses) and we argued that if we were able to show that all three of these approaches produced consistent results we would be able to make the clearest claims about the effects of the campaign. [Dr. Magura also expresses concern about the statistical analysis and the fact that we might have been reporting chance results; we dealt with this concern at some length in the paper and in the underlying technical reports and interested readers can review those documents (Hornik et al, 2010; Orwin et al 2006).]

4. We thus have specific responses to each of Dr. Magura's critiques

But it appears that the real difference between us is one of evaluation philosophy. Our evaluation is grounded in the recognition that in real life policy work no research will provide a definitive causal answer. It might even be argued that the act of imposing design control is almost guaranteed to distort the context of the intervention. We were asked to assess whether the campaign was affecting the youth of the United States as it was implemented; we were not asked whether a carefully controlled, unrealistically implemented program might work. Our judgment was that we would work as hard as we could to make strong and unbiased claims. Dr. Magura argues that if we could not address each of the concerns he raised, we should not have bothered at all. However our view comes from the opposite perspective: are the results we report and the inferences we make as strong as they can be given realistic resource and time constraints. Even as we recognize that every alternative interpretation has not been eliminated, are we successful in reducing uncertainty around policy decisions as far as it can be reduced, carefully balancing internal and external validity concerns. Is our work as good as we can make it for guiding policy – compared to the real alternative, making policy with no empirical basis whatsoever.

References

- Hornik, R., editor. Public Health Communication: Evidence for Behavior Change Mahwah. Lawrence Erlbaum; NJ: 2002.
- Hornik R, Jacobsohn L, Orwin R, Piesse A, Kalton G. Effects of the National Youth Anti-Drug Media Campaign on youths. *Am J Public Health*. 2008; 98(12):2229–36. [PubMed: 18923126]
- Orwin, R.; Cadell, D.; Chu, A.; Kalton, G.; Maklan, D.; Morin, C.; Piesse, A.; Sridharan, S.; Steele, D.; Taylor, K.; Tracy, E. Evaluation of the National Youth Anti-Drug Media Campaign : 2004 report of findings. Executive Office of the President, Office of National Drug Control Policy; Washington, D.C.: 2006. Available from <http://archives.drugabuse.gov/initiatives/westat/#reports>
- Palmgreen P, Donohew L, Lorch EP, Hoyle RH, Stephenson MT. Television campaigns and adolescent marijuana use: Tests of sensation seeking targeting. *Am J Public Health*. 2001; 91:292–296. [PubMed: 11211642]