



Highlights of GAO-10-30, a report to congressional requesters

Why GAO Did This Study

Recent congressional initiatives seek to focus funds for certain federal social programs on interventions for which randomized experiments show sizable, sustained benefits to participants or society. The private, nonprofit Coalition for Evidence-Based Policy undertook the Top Tier Evidence initiative to help federal programs identify interventions that meet this standard.

GAO was asked to examine (1) the validity and transparency of the Coalition's process, (2) how its process compared to that of six federally supported efforts to identify effective interventions, (3) the types of interventions best suited for assessment with randomized experiments, and (4) alternative rigorous methods used to assess effectiveness. GAO reviewed documents, observed the Coalition's advisory panel deliberate on interventions meeting its top tier standard, and reviewed other documents describing the processes the federally supported efforts had used. GAO reviewed the literature on evaluation methods and consulted experts on the use of randomized experiments.

The Coalition generally agreed with the findings. The Departments of Education and Health and Human Services provided technical comments on a draft of this report. The Department of Justice provided no comments.

What GAO Recommends

GAO makes no recommendations. View GAO-10-30 or key components. For more information, contact Nancy Kingsbury at (202) 512-2700 or kingsbyn@gao.gov.

November 2009

PROGRAM EVALUATION

A Variety of Rigorous Methods Can Help Identify Effective Interventions

What GAO Found

The Coalition's Top Tier Evidence initiative criteria for assessing evaluation quality conform to general social science research standards, but other features of its overall process differ from common practice for drawing conclusions about intervention effectiveness. The Top Tier initiative clearly describes how it identifies candidate interventions but is not as transparent about how it determines whether an intervention meets the top tier criteria. In the absence of detailed guidance, the panel defined sizable and sustained effects through case discussion. Over time, it increasingly obtained agreement on whether an intervention met the top tier criteria.

The major difference in rating study quality between the Top Tier and the six other initiatives examined is a product of the Top Tier standard as set out in certain legislative provisions: the other efforts accept well-designed, well-conducted, nonrandomized studies as credible evidence. The Top Tier initiative's choice of broad topics (such as early childhood interventions), emphasis on long-term effects, and use of narrow evidence criteria combine to provide limited information on what is effective in achieving specific outcomes. The panel recommended only 6 of 63 interventions reviewed as providing "sizeable, sustained effects on important outcomes." The other initiatives acknowledge a continuum of evidence credibility by reporting an intervention's effectiveness on a scale of high to low confidence.

The program evaluation literature generally agrees that well-conducted randomized experiments are best suited for assessing effectiveness when multiple causal influences create uncertainty about what caused results. However, they are often difficult, and sometimes impossible, to carry out. An evaluation must be able to control exposure to the intervention and ensure that treatment and control groups' experiences remain separate and distinct throughout the study.

Several rigorous alternatives to randomized experiments are considered appropriate for other situations: quasi-experimental comparison group studies, statistical analyses of observational data, and—in some circumstances—in-depth case studies. The credibility of their estimates of program effects relies on how well the studies' designs rule out competing causal explanations. Collecting additional data and targeting comparisons can help rule out other explanations.

GAO concludes that

- requiring evidence from randomized studies as sole proof of effectiveness will likely exclude many potentially effective and worthwhile practices;
- reliable assessments of evaluation results require research expertise but can be improved with detailed protocols and training;
- deciding to adopt an intervention involves other considerations in addition to effectiveness, such as cost and suitability to the local community; and
- improved evaluation quality would also help identify effective interventions.