




Considerations in Applying Benefit-Cost Analysis to Preventive Interventions for Children, Youth, and Families: Workshop Summary

ISBN
978-0-309-30105-3

82 pages
6 x 9
PAPERBACK (2014)

Steve Olson and Kimber Bogard, Rapporteurs; Board on Children, Youth, and Families; Institute of Medicine; National Research Council

 Add book to cart

 Find similar titles

 Share this PDF



Visit the National Academies Press online and register for...

- ✓ Instant access to free PDF downloads of titles from the
 - NATIONAL ACADEMY OF SCIENCES
 - NATIONAL ACADEMY OF ENGINEERING
 - INSTITUTE OF MEDICINE
 - NATIONAL RESEARCH COUNCIL
- ✓ 10% off print titles
- ✓ Custom notification of new releases in your field of interest
- ✓ Special offers and discounts

Distribution, posting, or copying of this PDF is strictly prohibited without written permission of the National Academies Press. Unless otherwise indicated, all materials in this PDF are copyrighted by the National Academy of Sciences. Request reprint permission for this book

CONSIDERATIONS IN APPLYING **BENEFIT-COST ANALYSIS**

**to Preventive Interventions
for Children, Youth,
and Families**

WORKSHOP SUMMARY

Steve Olson and Kimber Bogard, *Rapporteurs*

Board on Children, Youth, and Families

INSTITUTE OF MEDICINE *AND*
NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

NOTICE: The workshop that is the subject of this workshop summary was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

This activity was supported by Contract/Grant No.13-103067-000-USP between the National Academy of Sciences and the MacArthur Foundation. The views presented in this publication do not necessarily reflect the views of the organizations or agencies that provided support for the activity.

International Standard Book Number-13: 978-0-309-30105-3

International Standard Book Number-10: 0-309-30105-X

Additional copies of this workshop summary are available for sale from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; <http://www.nap.edu>.

For more information about the Institute of Medicine, visit the IOM home page at: www.iom.edu.

Copyright 2014 by the National Academy of Sciences. All rights reserved.

Printed in the United States of America

The serpent has been a symbol of long life, healing, and knowledge among almost all cultures and religions since the beginning of recorded history. The serpent adopted as a logotype by the Institute of Medicine is a relief carving from ancient Greece, now held by the Staatliche Museen in Berlin.

Suggested citation: IOM (Institute of Medicine) and NRC (National Research Council). 2014. *Considerations in applying benefit-cost analysis to preventive interventions for children, youth, and families: Workshop summary*. Washington, DC: The National Academies Press.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. C. D. Mote, Jr., is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. C. D. Mote, Jr., are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

PLANNING COMMITTEE ON STANDARDS FOR BENEFIT-COST ANALYSIS OF PREVENTIVE INTERVENTIONS FOR CHILDREN, YOUTH, AND FAMILIES¹

JEANNE BROOKS-GUNN (*Chair*), Virginia & Leonard Marx
Professor of Child Development, Columbia University
ANIRBAN BASU, Associate Professor and Director, University of
Washington, Seattle
JANET CURRIE, Henry Putman Professor of Economics and Public
Affairs, Princeton University School of Public and International
Affairs
JORGE DELVA, Professor and Associate Dean for Research, School of
Social Work, University of Michigan
ROSEANNE FLORES, Associate Professor of Psychology, Hunter
College, City University of New York
J. DAVID HAWKINS, Endowed Professor of Prevention, School of
Social Work, University of Washington, Seattle
LYNN KAROLY, Senior Economist, RAND Corporation
MELANIE LUTENBACHER, Associate Professor of Nursing and
Medicine, Vanderbilt University
DAN ROSENBAUM, Senior Economist, Economic Policy Division,
U.S. Office of Management and Budget
GARY VANLANDINGHAM, Director, Pew Charitable Trusts

Project Staff

KIMBER BOGARD, Project Director (*from November 2013*)
JOSHUA JOSEPH, Project Director (*until November 2013*)
WENDY KEENAN, Program Associate
DOUGLAS KANOVSKY, Senior Program Assistant

¹ Institute of Medicine planning committees are solely responsible for organizing the workshop, identifying topics, and choosing speakers. The responsibility for the published workshop summary rests with the workshop rapporteurs and the institution.

Board on Children, Youth, and Families Staff

FAYE HILLMAN, Financial Associate

PAMELLA ATAYI, Administrative Assistant

KIMBER BOGARD, Director

REVIEWERS

This workshop summary has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the National Research Council's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published workshop summary as sound as possible and to ensure that the workshop summary meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the process. We wish to thank the following individuals for their review of this workshop summary:

Anirban Basu, University of Washington
Lynn A. Karoly, RAND Corporation
Irwin Sandler, Arizona State University
David L. Weimer, University of Wisconsin

Although the reviewers listed above have provided many constructive comments and suggestions, they did not see the final draft of the workshop summary before its release. The review of this workshop summary was overseen by **Hugh H. Tilson**, University of North Carolina at Chapel Hill. Appointed by the Institute of Medicine, he was responsible for making certain that an independent examination of this workshop summary was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this workshop summary rests entirely with the rapporteurs and the institution.

Contents

1 INTRODUCTION AND THEMES OF THE WORKSHOP	1
Themes of the Workshop, 4	
Organization of This Report, 8	
2 BENEFIT-COST ANALYSES: EXAMPLES FROM THE FIELD	9
The Washington State Institute for Public Policy, 9	
Communities That Care, 14	
Evaluations by MDRC, 18	
3 ASSESSING THE COSTS AND BENEFITS OF INTERVENTIONS	21
An Ingredients Approach to Costing Preventive Interventions, 21	
Cost Analysis for Planning Purposes, 24	
Valuing the Outcomes of Intervention, 28	
4 ISSUES TO CONSIDER IN BENEFIT-COST ANALYSIS	35
Identifying Causal Estimates by Research Design, 35	
Designing Error-Tolerant Studies, 38	
Deciding What Evidence to Include, 40	
Issues with Randomized Controlled Trial Designs, 43	
Increasing the Comparability of Benefit-Cost Analyses, 44	
Expressing Uncertainty in Benefit-Cost Analyses, 48	
The Potential Compensation Test and Discount Rates, 49	

5 TRANSLATING RESULTS TO INFORM POLICY AND PRACTICE	53
A Perspective from the Office of Management and Budget, 53	
A State Policy Perspective, 55	
A Perspective from the Department of Health and Human Services, 58	
A Perspective from the States, 60	
REFERENCES	63
APPENDIXES	
A GLOSSARY	67
B WORKSHOP AGENDA	69

1

Introduction and Themes of the Workshop¹

Benefit-cost analyses hold great promise for influencing policies related to children, youth, and families. By comparing the costs of preventive interventions with the long-term benefits of those interventions, benefit-cost analysis provides a tool for determining what kinds of investments have the greatest potential to reduce the physical, mental, and behavioral health problems of young people (NRC and IOM, 2009). More generally, the growth of benefit-cost analysis as a field of research and practice represents an exciting and promising trend in the development and implementation of public policies.

The application of benefit-cost analyses to the field of prevention has been expanding rapidly. Preventive interventions occur prior to the onset of a disorder and are intended to prevent or reduce risk for the disorder (NRC and IOM, 2009). Benefits from investing in certain early childhood programs for economically disadvantaged children are among the best documented (Karoly, 2012), but benefit-cost analysis has been applied in many other areas as well (Lee et al., 2012; Pew-MacArthur Results First Initiative, 2012). For example, a 2006 benefit-cost analysis performed by the Washington State Institute for Public Policy demonstrated that by investing in a portfolio of evidence-based crime prevention programs, the Washington State legislature could reduce crime rates, avoid the need to construct a new prison, and save taxpayers \$2 billion

¹ The planning committee's role was limited to planning the workshop, and the workshop summary has been prepared by the workshop rapporteurs as a factual summary of what occurred at the workshop. Statements, recommendations, and opinions expressed are those of individual presenters and participants, and are not necessarily endorsed or verified by the Institute of Medicine or the National Research Council, and they should not be construed as reflecting any group consensus.

(Aos et al., 2006). In 2007, the legislature used these findings as the basis for expanding investments in evidence-based crime prevention, which resulted in a lowering of the state's long-term prison forecast such that a new 2,000-bed prison was no longer needed (Drake, 2010).

However, the utility of benefit-cost analyses has been limited by a lack of uniformity in the methods and assumptions underlying these studies. Researchers use a variety of techniques to calculate the costs of a program and the benefits it produces. They apply different discount rates to assign value to future costs and benefits. They report their results using different formats, with different levels of cost and benefit disaggregation and different levels of detail about underlying assumptions and uncertainties. For years, those who perform and those who use benefit-cost analyses have argued that the development and use of theoretical, technical, and reporting standards for benefit-cost analyses would enhance the validity of results, increase comparability across studies, and accelerate the progress of the field.

To explore this issue, the Board on Children, Youth, and Families of the Institute of Medicine (IOM) and the National Research Council (NRC) held a workshop on November 18–19, 2013, in Washington, DC, titled “Standards for Benefit-Cost Analysis of Preventive Interventions for Children, Youth, and Families.” The workshop constituted the first phase of a possible two-part effort directed toward guiding future benefit-cost studies and enhancing the relevance of benefit-cost analysis to governments and other organizations wanting to make sound prevention decisions. The workshop brought together leading practitioners in the field, researchers who study the methodological and analytic dimensions of benefit-cost analysis, and representatives of organizations that use the results of benefit-cost analyses to shape and implement public policies. Box 1-1 provides a list of questions that guided the development of the workshop’s agenda. A webcast of the workshop is available at <http://www.iom.edu/Activities/Children/AnalysisofPreventiveInterventions/2013-NOV-18.aspx>. This report summarizes the presentations and discussions at the workshop for researchers, practitioners, and policy makers and as input for a possible follow-on consensus study.

BOX 1-1
Statement of Task

An ad hoc committee will plan and conduct a 2-day public workshop to highlight the issues on finding consensus on the standards for benefit-cost analysis of preventive interventions for children, youth, and families. An individually authored workshop summary will be prepared based on the information gathered and the discussions held during the workshop sessions. The workshop will feature invited presentations and discussions that address the following questions:

- What level of research rigor should be met before results from an evaluation are used to estimate or predict outcomes in a benefit-cost analysis?
- What are best practices and methodologies for costing prevention interventions, including the assessment of full economic/opportunity costs?
- What prevention outcomes currently lend themselves to monetization? Are shadow prices available for those outcomes that are not typically monetized?
- What processes and methodologies should be used when theoretically and empirically linking prevention outcomes to avoided costs or increased revenues?
- Over what time period should the economic benefits of prevention interventions be projected and what discount rates should be used?
- What outcome domains are appropriate for a benefit-cost analysis of early childhood programs to consider?
- What are the best methods for handling risk and uncertainty in estimates? (E.g., What are the strengths and limitations of Monte Carlo simulations?)
- What information needs to be included in benefit-cost analysis summaries and reports?
- What issues arise when the results of benefit-cost analyses are applied to prevention efforts at scale? Do benefit-cost results from efficacy trials need to be adjusted when prevention is taken to scale?
- How should we account for heterogeneity in program effects in benefit-cost analyses?
- Can we define standards all studies should meet before they can be used to inform policy and budget decisions?
- How could research be used to create policy models that can help inform policy and budget decisions, analogous to the benefit-cost model developed by the Washington State Institute for Public Policy?
- What is the role of meta-analysis in the application of benefit-cost analysis to prevention programs?

THEMES OF THE WORKSHOP

In the final session of the workshop, the moderators of the preceding sessions summarized the major messages that emerged from the presentations and discussions that occurred during those sessions. Those messages are compiled in this section as an introduction to the themes of the workshop. These observations should not be seen as consensus recommendations of the workshop. Rather, they are points made by individual speakers that structured their presentations and the subsequent discussions.

The session moderators who participated in the final panel discussion were Jeanne Brooks-Gunn, the Virginia and Leonard Marx Professor of Child Development at Teachers College and the College of Physicians and Surgeons at Columbia University, who also chaired the planning committee for the workshop; Janet Currie, the Henry Putnam Professor of Economics and Public Affairs at Princeton University; Jorge Delva, professor of social work and associate dean for research in the School of Social Work at the University of Michigan; Roseanne Flores, associate professor in the Department of Psychology at Hunter College of the City University of New York; J. David Hawkins, endowed professor of prevention and founding director of the Social Development Research Group at the University of Washington School of Social Work; Melanie Lutenbacher, associate professor of nursing and medicine at Vanderbilt University; and Gary VanLandingham, director of the Results First Initiative, a joint project of the Pew Charitable Trusts and the John D. and Catherine T. MacArthur Foundation. Additional comments were made in the final session by Max Crowley, a research fellow at Duke University; Lynn Karoly, senior economist at the RAND Corporation; and Rebecca Maynard, University Trustee Chair Professor of Education and Social Policy at the University of Pennsylvania.

Standards in the Field

There has been significant progress in the field of benefit-cost analysis in recent decades. As Hawkins noted, even in 1980 very few such analyses had been done, and their effect on policy was negligible. But an increasing range of topics and program areas has been subjected to benefit-cost analysis, including preschool education, substance use pre-

vention, foster care, crime prevention, and health care (Barnett, 1985; Miller and Hendrie, 2008; Zerbe et al., 2009; Jacobsen, 2013).

The field of benefit-cost analysis could benefit from the development of standards in many different areas, said both Brooks-Gunn and Currie, including assessment of the cost of interventions, assigning values to outcomes, the use of randomized controlled trials and other experimental designs, applying discount rates, incorporating uncertainty into results, reconciling approaches across clearinghouses, and translating research results into a format useable by policy makers.

As Flores pointed out, a particular need is for better ways of monetizing the outcomes of interventions. For example, social and emotional development and other noncognitive skills can be important outcomes of an intervention, but what do those things mean and what is their economic value? As another example, standards on valuing the use of volunteer time could greatly facilitate comparisons across studies.

Reporting only statistically significant findings can leave out benefits and may have an effect on benefit-cost ratios, Hawkins observed. In addition, serendipitous secondary findings can lead to new research avenues that advance the field. One way to help standardize the reporting of outcomes would be to provide baseline information along with the uncertainties associated with that information.

Several of the panelists made the point that the costs and benefits of interventions vary across groups, locations, and times, which complicates both the analysis and replication of interventions. However, Delva added, new electronic technologies can capture data more quickly and accurately than in the past and may be a way to overcome these obstacles.

Research Designs

Randomized controlled experiments can provide valuable information about preventive interventions, several panelists said, but such trials are not always feasible, and standards for randomized controlled trials would be very helpful. VanLandingham added that a thorough description of the control groups in randomized controlled trials can help people understand whether they will be able to replicate a program in their setting and what the key elements of an intervention are.

Research designs other than randomized controlled trials can be appropriate and useful, Currie said. Different designs will have different

standards, but all research studies can be well or poorly designed. Criteria could be established that different kinds of designs need to meet, Brooks-Gunn suggested. Many of these issues are being considered in the area of public health and health care, and cross-fertilization among the fields could yield progress. Karoly pointed to the value of administrative data both in short-term evaluations and in learning about long-term impacts.

Incorporating an ethnographic component into benefit-cost analyses could increase understanding of what a program is doing and how it differs from other programs, Delva pointed out. The application of standards to qualitative research also could help increase comparability across studies, added Flores.

Clearinghouses to Disseminate the Results of Benefit-Cost Analyses

Clearinghouses can play a critical role in collecting and disseminating information, but, as VanLandingham observed, greater uniformity in the formats used to gather, analyze, and report data could make results more useable. Clearinghouses also could serve a useful function by providing all the information that may be of value to policy makers, not just the positive results.

Though the clearinghouses were set up for purposes other than standardization, they could be adapted to a common framework, said Maynard—a step currently being considered by a federal interagency group. In addition, existing clearinghouses can be used to examine and deliberate over standards of rigor. In particular, experimental designs other than randomized controlled trials could be considered so policy makers can take advantage of the full range of information that is available. At the same time, clearinghouses could be given the flexibility to address other questions and perspectives and to be flexible in the use of data.

The Development of Standards

Standardization is a process, Currie emphasized. Consensus may exist in some areas, while in other areas consensus will need to be devel-

oped for standardization. Transparency and consistency in the development of standards can help obtain their acceptance.

Even where it is not possible to agree on standards, it may be possible to agree on principles that can guide decisions and that may, with time, lead to standards, said Lutenbacher. In addition, establishing standards for benefit-cost analyses provides an opportunity to educate people about the field and about how best to use the results of this research.

Eagerness for Results

With tight budgets and demands for accountability, policy makers and others are eager for information about which policies work, which policies do not work, and which interventions are cost beneficial. Policy makers also can enable benefit-cost analyses when they are developing and authorizing programs, noted Flores, especially if they are engaged in communication with researchers.

The continued development of benefit-cost analysis provides an opportunity to think differently about how government operates, Hawkins observed. Most government social programs have been organized around responding to a problem instead of preventing that problem. But the potential to study the effectiveness of preventive programs creates hope about avoiding health and behavior problems before they occur.

As VanLandingham said, policy makers will always have to make compromises in funding and setting up these programs, so they need to know which aspects of a program are critical. Programs have different components, and these components can be treated separately to build the most efficient and effective program possible, just as resources can be allocated to portfolios of programs in the most efficient manner. As Crowley added, costs then can be linked to program components to understand what resources need to be invested in what strategies and how those strategies lead to specific outcomes (Crowley et al., 2012).

Finally, whether a program is implemented with fidelity can have a major influence on whether it produces benefits and on the extent of those benefits, Hawkins reminded the group. Support for high-quality implementation may therefore be necessary to replicate a program's success. Otherwise, good programs may fail, leading policy makers to shun programs that have the potential to be successful.

ORGANIZATION OF THIS REPORT

The next chapter of this workshop summary describes several benefit-cost analyses of programs that have had significant effects on public policy. Chapter 3 turns to the costing of interventions and the economic assessment of a program's effects. Chapter 4 looks at several technical issues that arise in benefit-cost analyses, including the validity of research designs other than randomized controlled trials, the treatment of uncertainty, and discount rates. Chapter 5 considers benefit-cost analyses from the perspectives of several users of the results of those analyses. Finally, Appendix A includes a glossary of terms used in this summary, and Appendix B is the agenda from the workshop.

2

Benefit-Cost Analyses: Examples from the Field

Three speakers at the workshop provided compelling examples of the use of benefit-cost analyses to inform policy decisions. Though the examples are quite different, they reveal many of the issues that arise in gathering, analyzing, and disseminating benefit and cost data. They also demonstrate both the opportunities and the challenges of creating greater standardization in the field.

THE WASHINGTON STATE INSTITUTE
FOR PUBLIC POLICY¹

In 1983 the Washington State legislature created the Washington State Institute for Public Policy to carry out practical, nonpartisan research at the direction of the legislature or the institute’s board of directors. A major activity of the institute is to conduct benefit-cost analyses of policy changes being considered by the state. As the institute’s director, Steve Aos, said at the workshop, the institute essentially functions as an investment advisor for the spending authority of government. It produces “buy and sell information” that legislators can use to make policy decisions.

The institute has looked at the benefits and costs of a very broad array of policies, including those affecting

- crime,
- education and early education,
- child abuse and neglect,

¹ This section summarizes information presented by Steve Aos, M.S., Washington State Institute for Public Policy, Olympia, Washington.

- substance abuse,
- mental health,
- health care,
- developmental disabilities,
- teen births,
- employment and workforce training,
- public assistance,
- public health, and
- housing.

The institute goes through a three-step process to determine benefits and costs. First, it examines what works to improve outcomes and what does not work. It applies a meta-analytic approach to all of the rigorous evaluations that it can identify of policies designed to improve public outcomes of legislative interest. It then combines studies that would be relevant to the state of Washington, a procedure AOs described as not only good science but good politics. Legislators can be suspicious of results based on a single study, he said, but if the results derive from every relevant study done in a particular field, those results have resonance.

Next, the institute examines the return on investment by computing the benefits, costs, and risks to the people of Washington State of a policy change (Lee et al., 2012). It uses a consistent framework to enable comparisons among options. This framework has developed over time, starting from a simple spreadsheet and evolving into a much more comprehensive evaluation.

Finally, the institute uses this information to help form budgets by exploring how a combination of options would affect statewide outcomes. This portfolio approach raises the analytic bar by considering such issues as diminishing returns when the beneficiaries of a policy are affected by multiple programs. All three steps are necessary, said AOs. In particular, the development of a portfolio of programs is just as much in need of standards as the evaluation of costs and benefits.

The product of this three-step process is a list of evidence-based policy options ranked by return on investment. The institute tries to present these in a *Consumer Reports* style, so the results look the same for a legislator working on K–12 education as for a legislator working on the juvenile justice system. Over the years, legislators have become accustomed to the format of the presentation and to being able to compare

across programs. “That has proven to be very fruitful across a wide number of areas,” Aos said.

The Nurse–Family Partnership as an Example

Aos looked in depth at the Nurse–Family Partnership program as an example of the institute’s approach. The costs of the program are derived from the costs of hiring nurses in Washington State labor markets, with the costs of instruction, service, and other activities included. Outcomes include reduced child welfare and victim costs, lower criminal justice costs, and savings in public assistance and health care. For example, the lower child welfare costs take into account the proportion of cases placed out of their homes and the marginal costs for foster care and other services provided to these youth.

Evaluations of the Nurse–Family Partnership program have found that educational attainment increases both for the children and mothers in the program, Aos stated. This increases K–12 costs as these individuals receive more schooling, including special education services. The reduced criminal justice costs are based on a detailed model that tracks every step from arrest through incarceration to community supervision upon release. According to the institute’s most recent results, the program produces a net benefit of almost \$17,000 per family (see Table 2-1). The benefit to cost ratio is 2.73, and the return on investment is 8 percent.

Accounting for Uncertainty

Each of the estimated benefits and many of the costs have a standard error. The institute uses a Monte Carlo simulation with these errors to calculate the likelihood that a program will be beneficial in any individual case. For example, with the Nurse–Family Partnership, the net benefits are positive 76 percent of the time and negative 24 percent of the time. Though this is more risk that many people associate with the program, it still indicates the program is a solid investment, Aos observed.

TABLE 2-1 Return on Investment for the Nurse–Family Partnership in Washington State

Benefits per Family		Main Source of Benefits
Reduced child abuse and neglect	\$1,096	Lower CW & victim costs
Increased ed. attainment (child & mother)	\$24,131	Increased earnings
Reduced crime (child & mother)	\$5,333	Lower CJ & victim costs
Increased K–12 costs	–\$1,738	Higher K–12 costs
Other	\$2,854	Pub asst, health care \$
Deadweight cost of program	–\$4,933	
Total Benefits per Family	\$26,743	
Cost per Family	\$9,788	
Net Benefits (NPV)	\$16,956	= \$2.73 B/C = 8% ROI

NOTE: B/C = benefit/cost; CJ = criminal justice; CW = child welfare; NPV = net present value; ROI = return on investment.

SOURCE: Aos, 2013.

The analysis of the Nurse–Family Partnership program still could be improved, Aos said. He is eager to see more evaluations of the program, especially in situations comparable to those in Washington State and by evaluators who are not associated with the design or delivery of the program. Replications of programs may not be as effective as its initial implementation, which suggests that it may be necessary to discount results of evaluations done by a program’s developers.

Lessons Learned

Aos drew several broad conclusions from the experiences of the Washington State Institute for Public Policy. First, the results of benefit-cost analyses need to compare apples to apples, not apples to oranges. Aos tries to give legislators several options to meet a particular policy goal. So long as they are evaluated the same way, legislators can make decisions based on consistent evidence.

Second, results have to be understandable by all 147 members of the Washington State legislature—or, said Aos, at least by the committee chairs, the majority leader, and the ranking members. For that reason, methods need to be both intuitive and scientifically justifiable. Legislators also need to know about the uncertainty of results. If results cannot be explained in an understandable way to legislators, they might get published in a journal, but they will not affect policy.

Third, the institute's results are calculated on an annual cash flow basis from three perspectives: that of taxpayers, that of participants in the program, and that of others who are affected by the program, such as the victims of crime. Different legislators can be interested in different aspects of the results, Aos pointed out, depending on whether they are on the fiscal committee, for instance, or the juvenile justice committee.

Fourth, the effect size of a program is important, but so is the risk associated with that estimate, said Aos. Legislators need to know about the uncertainty of results from benefit-cost analyses. Aos also warned against the tendency to double-count benefits. For example, some outcomes measure the same human capital construct, such as higher test scores and increased high school graduation. The institute uses trumping procedures to avoid constructs that end up measuring the same thing.

Fifth, because local conditions vary in the United States, the results of benefit-cost analyses will, too, indicated Aos. The same program that reduces crime in Texas will save more taxpayer money than it will in Washington State where fewer people go to prison. State-specific numbers are needed to reflect local conditions. In addition, results and methods need to be updated at least annually. These ongoing updates can incorporate new studies, model refinements, and better data.

Sixth, Aos recommended greater use of longitudinal research to estimate benefits and costs. For example, child abuse and neglect are linked to later outcomes such as a failure to graduate from high school or increases in crime. Longitudinal research using large datasets is now drawing these linkages, which need to be incorporated into the results of benefit-cost analyses.

Finally, Aos urged practitioners to borrow the best current thinking on the valuation of outcomes. In this way, work does not need to be redone, but it may need to be adapted to local circumstances.

The state budget in Washington State has been affected by the institute's results, said Aos. As an example, he mentioned forecasts of the number of prison beds that will be needed in the state, which draw on

results from the institute (Aos et al., 2006). Uptake has been faster in some areas than others, but it is gratifying to be perceived as an honest broker and to supply information that makes a difference.

COMMUNITIES THAT CARE²

Communities That Care is a preventive intervention that takes a public health approach to promoting positive youth development and reducing problem behavior. As described by Margaret Kuklinski, Communities That Care relies on coalitions of diverse stakeholders, including mayors, police chiefs, teachers, and parents, who receive training and carry out Communities That Care in their communities. The intervention begins with a survey of the youth in a community to understand where risk factors are elevated and where protective factors are depressed. The coalition then selects and implements evidence-based prevention programs to reduce the most widespread elevated risks. The coalition monitors fidelity of implementation, assesses efficacy, and makes course corrections when needed to achieve the community's overall prevention goals, with a coordinator overseeing these diverse activities.

To determine the efficacy of the intervention, a randomized controlled trial was conducted involving 4,407 students in 24 communities and 7 states. The communities were matched in pairs within states and then randomized to condition, with no significant differences at baseline on important sociodemographic characteristics. Students were followed annually for 10 years, from fifth grade through age 19, with more than 90 percent of youth being followed over that period. Youth exposed to the program reported significantly lower rates of initiation community wide with respect to delinquency, alcohol use, and cigarette smoking, Kuklinski reported. Youth showed nonsignificant results for high school graduation, marijuana use, and other drug use initiation, but the observed changes were in the expected direction.

The program invested an average of \$745,000, measured in discounted 2011 dollars, in each community over the course of 5 years in training, technical assistance, monitoring, preventive programs, and coordination. However, the costs varied across the 12 communities in the

² This section summarizes information presented by Margaret Kuklinski, Ph.D., University of Washington, Seattle.

study, Kuklinski observed, from \$283 per youth to a high of \$5,730 per youth, partly because of the differing sizes of the communities involved in the program and economies of scale in the implementation of Communities That Care. In deciding on a single estimate for the cost per youth, evaluators initially used an average cost of \$1,159 as providing a conservative estimate. Later they determined that a weighted average of \$556 per youth was a better estimate of the cost in communities in the sample as well as those likely to implement the program. They also used a range of 35 percent around that point estimate in Monte Carlo analyses to account for the variability in cost.

The Communities That Care program relied heavily on board members, volunteers, and teachers donating their time, which raises additional questions about how to value these nonbudgetary economic resources, Kuklinski pointed out. One approach would be to assume that the time cost is fully offset by the benefits volunteers and teachers receive, in which case the net opportunity costs would be zero. A second approach is to value that time at the appropriate wage rate plus a fringe benefits rate for that position. A third option would value the time at the volunteer's own wage rate plus a fringe benefits rate. The evaluators found the first and second options most compelling. This meant that the opportunity costs could range from zero to as high as an additional \$89 per youth on top of the weighted average cost estimate.

In calculating benefits, Kuklinski continued, evaluators had to decide whether to monetize significant intervention effects only or all effects and what discount rate to apply to the benefits projections. If only significant effects were considered, the total benefits were \$4,477 per youth at a discount rate of 3.5 percent (see Figure 2-1). At a 7 percent discount rate, the value was nearly halved to \$2,312. If all effects were considered, the benefits were \$8,123 and \$4,070 at the respective discount rates, though the confidence intervals were greater around the point estimates when all benefits were considered.

Kuklinski showed that different viable assumptions lead to a range of conclusions. At the low end of assumptions, the benefit-cost ratio is 3.58, while at the high end of the assumptions, the ratio is 14.70 (see Table 2-2). In its analysis, the Social Development Research Group opted to include only significant effects, assumed no opportunity costs incurred for volunteers, and used a discount rate of 3.5 percent, arriving at a benefit-cost ratio of 8.22.

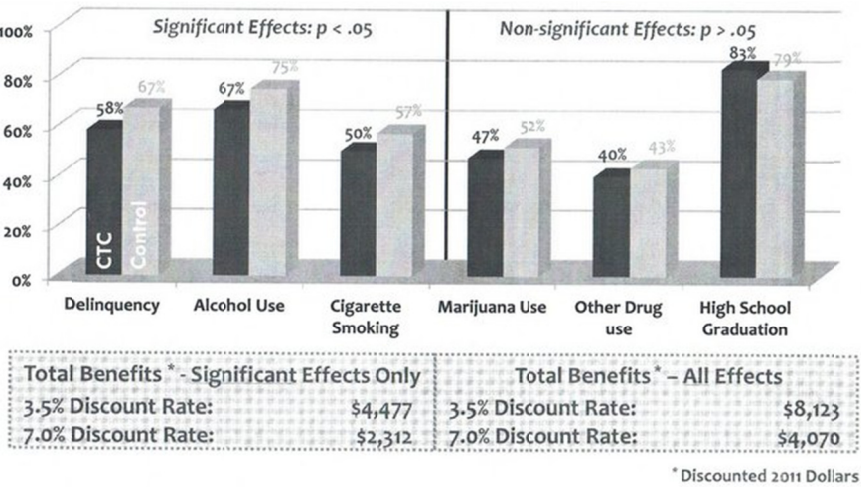


FIGURE 2-1 Including all effects or only significant effects and varying the discount rate have a substantial effect on the estimated benefits of the Communities That Care intervention.
SOURCE: Kuklinski, 2013.

Lessons Learned

Kuklinski drew several conclusions from the benefit-cost analysis conducted on Communities That Care. First, several important decisions have major implications for the bottom line of any such analysis. In the case of Communities That Care, the benefits included, the opportunity costs for volunteers, and the discount rate all had a substantial influence on the calculated benefit-cost ratio.

Standards for making some of these decisions would increase the value of findings in policy formation and improve comparability across studies, Kuklinski said. These standards could apply to research design, the assessment of costs, and the estimation of benefits. For example, in complex multisite trials where costs vary, the assessment of costs has implications for disseminating preventive interventions. Similarly, in some cases, it may be useful to specify different sets of outcomes to be included in benefit-cost evaluations within a specific program area. The logic models behind interventions can point to the important

TABLE 2-2 Different Viable Assumptions Lead to a Range of Benefit-Cost Ratios for the Communities That Care Intervention

	Range of Estimates*		SDRG Analysis
	Low	High	
Assumptions	Significant Effects, Opportunity Costs = \$89, 7% Discount Rate	All Effects, Opportunity Costs = \$0, 3.5% Discount Rate	Significant Effects, Opportunity Costs = \$0, 3.5% Discount Rate
Benefits	\$2,312	\$8,123	\$4,477
Costs	(635)	(556)	(556)
Net Present Value	\$1,667	\$7,617	\$3,920
Benefit-Cost Ratio	3.58	14.70	8.22

* Discounted 2011 dollars.

SOURCE: Kuklinsky, 2013.

outcomes to assess rather than focusing on the most broad set of outcomes to be monetized.

The field has not been clear about best practices, Kuklinski said, which raises the need for standards. Moreover, the need for standards has increased as the number of studies and applications has increased. The results of benefit-cost analyses can include many types of information, including internal rates of return, cash flows, investment risks, discount rates, and benefits and costs organized by various stakeholders. Adherence to best reporting practices may be needed for the users of information to have confidence in what they are reading.

A challenge for a future study of standardization in benefit-cost analyses would be to identify areas where common ground exists. For example, the appropriate discount rate, or a range of rates, may be an area where consensus could be achieved. Where consensus is less clear, the major alternatives need to be delineated. Greater standardization “would be of great use to researchers like me, to policy makers, and to practitioners who use this work,” Kuklinski said. “It would increase methodological consistency and help us be able to make meaningful comparisons of results from different studies.”

EVALUATIONS BY MDRC³

Charles Michalopoulos described several evaluations that MDRC has done of preventive interventions. First, the Minnesota Family Investment Program is a welfare-to-work program that provides work incentives to welfare recipients in Minnesota and a range of supports such as child care. Study of the program found that it increased employment and income among participants. At the same time, evaluators found that the program had benefits for children whose parents were in the program. In particular, children's academic achievement rose, especially among young children. The program cost the government about \$2,000 per participant per year. However, the evaluations did not monetize all of the benefits, such as improvements in child well-being, increases in the rate of marriage among program participants, and distributional effects. Also, the analysis did not present a measure of uncertainty.

A second example is a study of the Foundations of Learning program, which is an intervention to help Head Start teachers in their classroom management and to provide classroom mental health consultants. Results suggested that the program improved teachers' ability to manage their classrooms and reduced problem behavior, but it did not have other benefits for children once they entered school. In particular, it did not increase their mathematics and reading skills. The costs of the program were \$1,750 per child, but the benefits, such as reduced problem behavior, could not be easily monetized. In contrast, a goal such as avoiding special education, which is quite expensive, could be monetized, Michalopoulos observed.

Michalopoulos also discussed evaluations of programs based on the Nurse-Family Partnership model. Three randomized controlled trials of programs found savings in health care, welfare, and criminal justice costs. The Coalition for Evidence-Based Policy recently declared that the Nurse-Family Partnership program is in the first tier for being an effective program. The success of this program had a substantial effect on the Affordable Care Act, which allocated \$1.5 billion for implementing and studying home visiting programs. The program also is currently reimbursed through some state Medicaid programs.

³ This section summarizes information presented by Charles Michalopoulos, Ph.D., MDRC, Oakland, California.

Lessons Learned

Experience with these and other programs illustrates several challenges related to measuring costs, Michalopoulos said. First, one-time costs, such as teacher training in the Foundations of Learning program, can be problematic. Once teachers were trained, they could keep benefiting children as long as they stay in their jobs. Should the initial costs be spread over a period of time? This is an area where standards for benefit-cost analyses would be useful.

A related question relates to the differences between demonstration programs and ongoing programs. Again, with the Foundations of Learning project, teachers already undergo training in the course of their work. If new training could be added to existing training, the costs of administering the program would be reduced.

Many benefit-cost analyses do not devote enough effort to express statistical uncertainties about the results, Michalopoulos said. In a study of benefit-cost analyses for early childhood interventions, Karoly (2012) found only three studies that expressed statistical uncertainty. However, large uncertainties may lead to different policy implications than would small uncertainties.

Michalopoulos also made the point that significance tests are often the wrong choice for making policy decisions. For example, an intervention with an average net benefit of \$400 and a standard error of \$200 would be statistically significant, but if the standard error were \$300, it would not be significant. Yet the second intervention would still have a 91 percent chance of saving money. In addition, an outcome that is not statistically significant can be substantially more important than an outcome with an equal confidence interval. A better measure than statistical significance, said Michalopoulos, would be whether an intervention is likely to save money or meet some other objective.

Including results that are not statistically significant points to the advantages of designing a benefit-cost analysis ahead of time rather than looking at the results of an analysis and deciding which to include in an analysis, Michalopoulos pointed out. However, it may be necessary to omit measures that have large amounts of uncertainty. For example, projections 20 years into the future may be too uncertain to include in a study. Adding a benefit with great uncertainty to more certain benefits may create too much uncertainty for the overall measure of benefits,

reducing the value of a study. Also, some outcomes may be too difficult to monetize, even if they seem important.

These kinds of decisions can be difficult to make Michalopoulos admitted. For example, outcomes that occur in the near term are often more certain than long-term outcomes. As another example, the indirect effects of an intervention on peers of siblings may be important but hard to measure. MDRC studies typically focus on outcomes that are relatively certain while pointing to other factors that may make outcomes better or worse.

Long-term follow-up can reduce uncertainties in benefit-cost analyses, Michalopoulos observed. However, the question has to be asked whether outcomes observed in one context and with one group can be generalized to other contexts or groups. One way to answer this question is simply to assume the benefits will be lower in a different setting; but how much lower? Multisite and multisubgroup studies can help answer these questions, but results may change with the assumptions made. Also, circumstances can change over the course of an extended study and increase uncertainty in overall outcomes. Focusing on a few outcomes with greater certainty can be supplemented by longer-term projections.

Michalopoulos urged against developing standards that exclude good research or codify bad research. For that reason, establishing guidance or principles may be a better approach than defining standards. He also called attention to two areas of tension: the contrast between the complexity of benefit-cost analyses and the simple answers policy makers want and need, and the need to focus on key outcomes while still measuring everything of importance.

The diversity of outcomes from most interventions emphasizes the need to focus on key outcomes that can be monetized with reasonable certainty, Michalopoulos concluded. Many relevant benefits can be measured, but which benefits to monetize remains an important question. For example, the target age of a child may be less important than whether the benefits are monetizable. Also, the intended uses of information can affect the outcomes being measured. For example, a department of health may be more worried about child health outcomes while a department of human services is more worried about cognitive development. However, in comparing programs with similar goals, application of the same standards would be very useful.

3

Assessing the Costs and Benefits of Interventions

At the heart of conducting a benefit-cost analysis is assessment of the benefits and costs of an intervention. Both of these dimensions are challenging, but could benefit from greater standardization in the field. Two speakers at the workshop looked specifically at the costing of interventions, while one explored the valuing of benefits.

AN INGREDIENTS APPROACH TO COSTING PREVENTIVE INTERVENTIONS¹

Benefit-cost analyses tend to use calipers to measure effects and witching rods to measure costs, said Henry Levin. In many cases, costs are all but ignored because analysts are so focused on finding effectiveness results. This includes not just overt cost transactions but other costs that together represent the true costs of an intervention. For example, a budget is not necessarily a full or accurate metric for determining costs, Levin observed. Nor do administered prices in a hospital or grant support for a community intervention provide accurate pictures of the resources required to produce benefits.

In determining costs, important questions include

- What are the criteria for determining costs?
- How complete are the costs? Do they cover all of the requirements needed to produce the effects on which benefits are based?

¹ This section summarizes information presented by Henry Levin, Ph.D., Columbia University, New York, New York.

- Do the costs use comparable prices for comparison (e.g., local versus national prices for goods and services can vary dramatically in price)?
- Is the information adequate for an observer to replicate results?

Standards for benefit-cost analyses will not be easy to develop, Levin acknowledged, but they may help answer these questions.

The Ingredients Cost Method

Levin uses what he called the “ingredients cost method” to determine costs. It relies on the use of competitive market prices or shadow prices based on markets. Its goal is to ascertain the cost of all the resources required to replicate an effectiveness result.

The method has several major steps. First, for each alternative, an intervention and its theory of action need to be described. This enables a benefit-cost analysis to reflect what was being attempted and how and why outcomes occurred.

Second, the specific ingredients or resources used to implement the intervention need to be described in terms of both quantity and quality, irrespective of how they are financed, Levin said. A volunteer can be self-financed, but a volunteer is not simply free, because in another setting the market cost for that input must be paid. Documents, interviews, and observations can identify these ingredients or resources, although these sources of information may not be fully available for interventions conducted in the past.

Third, market or shadow prices are assigned to all ingredients based on opportunity costs. This is done independently of funding sources, including in-kind resources such as volunteers. Finally, the costs are analyzed in different ways to make them amenable for analysis and comparison. For example, marginal costs or average costs may be calculated. These costs can be presented in the form of worksheets that identify the categories of program ingredients and the constituencies who are paying for those ingredients (see Figure 3-1). Developed during the past four decades, this method has been computerized and includes a database of prices, discount rates, and other data, Levin said. For example, the Poverty Action Lab at the Massachusetts Institute of Technology has adopted this method to do cost-effectiveness studies of its randomized controlled trials.

Column 1: Cost Ingredients	Column 2: Total Cost	Column 3: Cost to Program Sponsor	Column 4: Cost to Other Government Agencies	Column 5: Cost to Other Private Organizations	Column 6: Cost to Students and Parents
Personnel					
...					
Facilities					
...					
Materials and equipment					
...					
Other inputs					
...					
Required client inputs					
...					
Total ingredients cost					
User fees		-()			+()
Cash subsidies		-()	+()	+()	
Net costs					

FIGURE 3-1 Worksheet identifies types of costs and the groups on which costs are imposed.
SOURCE: Levin, 1975. Reprinted with the permission of Sage Publications.

Levin particularly pointed out that the identification of costs needs to be done separately from a program’s financing. For example, over time, a program may not be able to sustain volunteer efforts, in which case it may be necessary to look for financing to cover those costs. Separating financing from costs is a bedrock principle of cost accounting.

Levin also pointed out that doing a cost analysis can sensitize people to the assumptions made in that analysis, just as with analyses of benefits. For example, a question that can arise is what does the time and effort that a student is putting into a program displace? That is not an easy question to answer. Some interventions may argue that they cost nothing or very little, but the ingredients cost method demonstrates that their cost may be fairly high. For example, some interventions simply reallocate resources from one program to another (e.g., from health prevention to educational enrichment) and argue that there is no cost. Of course, the cost is determined by the value of what is sacrificed by abandoning the initial program rather than assuming that such transfer is “free.” Figuring out what these opportunity costs are is important, said Levin.

A Call for a Civil Union

Levin urged setting standards for cost analysis that are equally rigorous as those for both effectiveness and benefits analysis. Costing should not be an afterthought, he said, but should be closely related to and done simultaneously with benefits analysis, with the sharing of data for a complete evaluation.

Analyses also need to incorporate a strong ethnographic component that documents the intervention process and ingredients, Levin added. Such analyses, which could be done parsimoniously with periodic visits and data gathering, can reveal what really happened during an intervention, not just what theory would predict should happen. Qualitative analyses also can help explain differences in site results.

Finally, retrospective cost analysis should be avoided, Levin said. Though sometimes necessary because costs were ignored at the time of the intervention, such analyses irretrievably lose much information. If done retrospectively, it is important for cost analyses to be as timely as possible.

COST ANALYSIS FOR PLANNING PURPOSES²

Whether with expanded home-visiting programs, school violence prevention efforts, or early learning initiatives, robust cost estimates are needed to support prevention efforts that can effectively meet public health needs and reduce the strain on overburdened service systems, said Max Crowley. In particular, by demonstrating the resources needed for prevention, cost analyses are inextricably linked to efforts to take interventions to scale. “Even basic cost estimates can end up having greater utility for program planning than some of our best estimates of program cost-effectiveness or benefit-cost ratios,” Crowley said.

Crowley discussed three key issues related to the process of costing prevention programs. The first involves the resources needed to ensure adequate programming infrastructure is in place. Many estimates of prevention costs capture only the most immediate resource needs of programs. In particular, they often neglect crucial elements of infrastructure.

² This section summarizes information presented by Max Crowley, Ph.D., Duke University, Raleigh, North Carolina.

For example, local knowledge about how to adopt and implement preventive programs can vary tremendously on the ground. Many programs now use manualized training to teach program facilitators how to deliver a specific prevention curriculum, but few programs seek to train the managers overseeing those facilitators to ensure programs are delivered as they were intended and with quality. Nor are these managers generally taught how to enable themselves to fend off threats to the sustainability of a prevention effort, even though such training is often crucial to replicating the effects of an evaluation trial. These skills are often assumed to be available in the existing labor market, but the reality is that these skills may be in very short supply, especially in rural or impoverished areas. To successfully deliver preventive programs and replicate the effectiveness of trials, this local capacity must be deliberately built through training and technical assistance, Crowley said, which can require significant resources. “If we don’t budget for infrastructure, we can undermine the whole prevention effort.”

Infrastructure building can be divided into three main areas: adoption capacity, implementation capacity, and sustainability capacity. Adoption capacity refers to the ability of a local community to attract or train a labor force with the ability to adopt an evidence-based prevention program. This involves local capacity to understand the needs of the target population as well as the fit of a program in that community (Lutenbacher et al., 2002). Prevention delivery and support systems such as the PROSPER (*PRO*moting *S*chool–community–university *P*artnerships to *E*nhance *R*esilience) Network and Communities That Care (see Chapter 2) seek to supplement this local capacity when it is lacking. These systems also allow for different programs to be deployed depending on local needs. Such “plug and play” systems allow program evaluators to include infrastructure development in their cost analyses with less research burden than if a program were implemented in isolation.

Implementation capacity refers to the ability to deliver the program from manualized curricula and to ensure program quality. Many communities lack existing quality assurance systems that are compatible with many of the prevention programs currently available in the marketplace. Developing these systems can take time, but they are essential to ensure prevention services are delivered with fidelity (Durlak and DuPre, 2008).

Sustainability capacity refers to the ability of a prevention effort to integrate a program into the existing service infrastructure and develop robust funding streams. Developing this capacity requires training and

ongoing technical assistance around fundraising and management of in-kind and volunteer resources.

Volunteer and In-Kind Donations

The next issue Crowley discussed involves volunteer and in-kind donations. Because they attempt to avert a future outcome, prevention programs often require substantially greater buy-in than do downstream solutions that seek to triage existing and visible problems (Elliott and Mihalic, 2004). As a result, prevention initiatives often rely heavily on local volunteer and in-kind donations. Securing these resources not only solidifies buy-in from the community but also can help alleviate resource scarcity (Feinberg et al., 2008).

Program evaluators often seek to estimate the cost of a program retrospectively (Crowley et al., 2012). But institutional knowledge of what group donated which resources and which people volunteered their time is often lost in such analyses. This is particularly a threat for programs that rely on existing service infrastructures to house their programming. While there will always be a place for retrospective analyses, Crowley acknowledged, the issue is planning for cost analyses up front in a project. By building an overall data architecture, costs can be captured in more naturalistic settings so the process is less burdensome.

As an example, Crowley noted that the education system has often been the natural home for preventive interventions targeting adolescent populations. The opportunity costs of such a program may be small when they are delivered within trials, but when they are delivered at scale, they often require sizable investments from local education systems. These investments include not only the more visible teacher time but also the less visible administrative and staff time. When these costs are not captured, they can threaten program planning and place an unexpected burden on service systems, possibly derailing an entire program.

Participant Costs

When prevention programs seek parental support, as is often the case with family-based programs, this support represents a cost, Crowley pointed out. Parents are also the targets of many preventive programs.

Thus, participant costs can include the cost to a child, the cost to a parent, and the cost to a family.

Participant costs require special handling, Crowley observed. Sometimes costs are incurred up front, often through losses in time, but sometimes they are incurred as part of programs themselves. For instance, preventive programs that improve parents' success in the labor market can produce a cost to children. The hope is that child participants will gain an overall benefit from the program. But the cost to a parent of not being able to provide child care can be important for recruitment and participation in a program and can inform program planning.

Self-report interviews can identify losses of service, but they can fail to capture the complexity of program costs, especially in the context of increasingly dynamic preventive efforts that are delivered across substantial periods of time. However, with the development of more robust data collection systems, particularly through new technological supports, the field can dramatically extend the science around prevention costs, Crowley observed. For example, the geographic information systems being deployed by many community prevention efforts can capture such costs as participant travel time to participate in a family-based program. Looking farther into the future, e-health technologies could collect large quantities of data to quantify participant costs.

Potential Best Practices

Crowley concluded by pointing to several best practices that can guide the development of standards for cost analyses. First, he said, cost analysis should always be prospective. Failing to include a cost analysis at the beginning of a trial makes the process of estimating costs much more difficult and increases the likelihood that cost estimates will be incomplete. Ideally, funders and reviewers will someday expect a program evaluation to include capturing opportunity costs.

Another best practice, Crowley observed is to use the ingredients-based approach described earlier by Levin to capture a full economic accounting of a program and deconstruct the resource needs into specific cost categories. Such an approach would yield more standardized protocols for cost collection and would link resource consumption to program activities.

Third, cost analyses should always seek to estimate the full economic costs of implementing the program, Crowley said. Cost estimates need to move beyond simple budgetary review and include all of the resources needed to replicate a program's effects. These resources include those needed to build local capacity, the cost of donating time and goods, and a full accounting of participant costs.

Fourth, Crowley pointed to the need to explore uncertainties in cost estimates. Sensitivity analyses could test the robustness of cost estimates under a variety of assumptions. Monte Carlo analyses can be applied to the costs of a program as well as to the benefits to yield confidence intervals around point estimates.

Crowley also identified four areas that could benefit from greater standardization:

1. Identify essential cost categories that all cost analyses should strive to include.
2. Develop guidelines for appropriate handling of costs that are not reflected in program budgets.
3. Establish minimum levels of sensitivity analysis to explore uncertainty in cost estimates.
4. Ensure consistent reporting of cost estimates to enhance transparency and utility.

Prevention programs, especially for children, youth, and families, are increasingly in the spotlight, Crowley observed. Well-done cost analyses can describe these investments and help communities decide which investments to make. But cost estimates need to capture all of the resources needed for a program to avoid jeopardizing the quality of the services being delivered and the sustainability of a program.

VALUING THE OUTCOMES OF INTERVENTION³

When a preventive intervention is evaluated, the headline is often the total economic return the program will generate. But this headline

³ This section summarizes information presented by Damon Jones, Ph.D., The Pennsylvania State University, University Park, Pennsylvania.

obscures a number of important questions about what is included in the total dollar amount, said Damon Jones, including

- Is the return based mostly on one sector?
- Does it involve a combination of many sectors?
- Who is receiving the economic benefits—participants, taxpayers, other nonparticipants?
- When does the savings or benefit occur?
- Is the benefit based on projections?

These are the kinds of questions that need to be addressed to measure the economic outcomes of an intervention, Jones stated.

Previous benefit-cost analyses of early childhood interventions have found benefits in a variety of economic and societal sectors. For example, evaluations of the Perry Preschool Project have found lower crime rates, less retention and special education use in school, and increased lifetime earnings (Belfield et al., 2006). Study of the Chicago Child Parent Centers revealed increased earning and tax revenues, reduced costs associated with crime, and reduced need for special education services (Reynolds et al., 2011). The Abecedarian Program demonstrated increased lifetime earnings for participants, increased maternal earnings, decreased school costs, and decreased smoking-related costs (Barnett and Masse, 2007).

An important question with these and other studies is how these outcomes were monetized, said Jones. For example, shadow prices for valuing interventions are often a critical element of this process. Shadow prices evaluate the return on such interventions as increasing high school graduation or test scores or reducing criminal acts or early substance use. Shadow prices are harder to determine for outcomes measured at younger ages, Jones observed. Shadow prices also are easier to apply with some sort of categorical result, such as a diagnosis. And the more projection required to determine a price, the more uncertainty needs to be accounted for in the methodology.

Standards for Valuing Benefits

Currently there are no established standards for valuing benefits, Jones noted, especially with preventive interventions in the behavioral

sciences. Instead, economic evaluations are carried out in widely varying ways with respect to approach, measures, outcomes, and the structure of the assessment. In thinking about how to monetize benefits, a decision has to be made between determining how to value key study outcomes that may be indirectly linked to dollar amounts and focusing on the extent to which a program affects true economic outcomes. The answer to this question depends largely on the nature of the program, the age of the participants, and the outcomes that are targeted. Studies of preventive programs for children and families, for example, vary widely in terms of what outcomes are monetized. Many outcomes are left out because of a lack of precedence for how to monetize them, such as mental health outcomes, social skills, child behavior, and parenting skills.

Great variation also surrounds how effects are projected into the future. Some studies look only at the low-hanging fruit, which may lead to greatly underestimating economic impacts. But the difficulties involved in measuring the harder-to-obtain outcomes can lead to errors in estimates. At the same time, evaluators need to consider the full set of possible outcomes (both observable and those projected to occur) to achieve adequate coverage of economic impact.

A critical step for any evaluation is to think about what is left out, Jones added. When interventions are evaluated for economic impact retrospectively, the lack of planned measures may lead to incomplete assessment of benefits. But this could also be partly a matter of the inability to capture costs within a set of measured outcomes.

What may be particularly underestimated is the value of factors linked to long-term personal success, such as the development of interpersonal and intrapersonal skills, said Jones. The complexity of the interrelationship among different factors—such as cognitive and noncognitive skills, as well as context of these factors—is not easily represented in assessments of economic benefits. Even when the importance of certain skills for long-term personal success is recognized, the challenge is determining how to capture those skills in measuring outcomes that are not currently included. Possible approaches can build on prior research, but a standard methodology for modeling more complex associations is unlikely.

Valuing Outcomes

Although there are no clear standards for carrying out economic evaluation in preventive interventions, certain steps could be taken to increase consistency across studies. For instance, the first step in developing standards for valuing outcomes is to plan ahead, said Jones. As with the evaluation of costs, the economic evaluation component needs to be considered before establishing plans for the overall program evaluation, he observed. This requires consulting with economists who can provide expertise on how to structure an evaluation. It also requires reviewing prior research to learn what has been done, with what population, and in what context, and also checking whether any standards already have been established for economic evaluation.

The second step is to consider the scope and reach of a program's effects. What outcome domains will be affected? How far over time will these effects extend? Who will be affected? These questions reflect the logic model of the intervention, and while it is possible to include too many different effects, more is preferable to less, Jones said.

The third step is to determine the best measures for economic evaluation based on the intervention's logic model, Jones continued. Measurable outcomes as well as program benefits that cannot be measured need to be identified. Prior research may have used measures that can be applied or used in models. Deciding how to represent the uncertainty in valuations is also part of this step.

The fourth step in the process is to assess what key program outcomes cannot be valued, Jones said. Evaluators then need to determine whether the evaluation should incorporate other methodologies to determine economic benefits for these outcomes. If the outcome is a primary variable in the program evaluation, should the evaluation incorporate a cost-effectiveness component? Current or future research may help determine the possible valuation of these outcomes so they can be incorporated into later retrospective assessments.

Examples of the Approach

As an example of identifying potential outcomes, Jones described a hypothetical middle school preventive intervention program aimed at improving social skills and decreasing substance use in adolescents. The

program was delivered in the sixth grade through a curriculum occurring 2 days per week. It involved components such as demonstrative video modules, journal writing, and role-playing activities. A pilot study indicated multiple program effects measured at posttest, including fewer class disruptions, lower rates of bullying, increased engagement in class, and lower rates of initiation of substance use.

In subsequent research the evaluators wanted to include an assessment of the program's economic impact, Jones continued. They planned to assess the full cost and resources needed to deliver the program. They also planned to follow participants into high school to assess longer-term effects of the program. Prior research provided common methods for valuing outcomes in school programs, which helped the researchers determine what measures to include for this age group at posttest and follow-up assessment. These measures included use of special education services, class grades, grade retention, reported substance use, and use of other school services (including disciplinary and counseling services).

For a middle school program, participants can provide outcomes that are more readily monetized than for other populations or programs, Jones said. For instance, academic achievement can be identified at these ages and followed through high school. Moreover, substance use, delinquency, and early involvement in the justice system all can be measured.

Still, Jones observed, several key questions need to be answered. Are only the outcomes listed above valued, some of which also involve direct costs? How much can the costs from effects on current outcomes be projected into the future? For example, should reduced early substance use be projected to reduced longer-term problems? Should improved academic achievement be projected to future earnings? How about the value of outcomes that are not easily monetized? How long should effects be followed? For example, if participants are followed into young adulthood, should things like high school completion, college experience, early employment, longer-term substance use patterns, and longer-term delinquency and criminal activity be assessed? Evaluators also need to assess whom else might be affected by a program, such as teachers, other educators, family members, or the broader society.

The effects of a program can be distinguished by recipient, time frame, and whether they can be monetized. These potential outcomes can be derived from the logic model for a program, Jones stated, including both effects that can be monetized and those that cannot.

Programs for Younger Children

Some programs are at a disadvantage for economic evaluation based on the nature of the processes involved, Jones said. For example, younger children usually cannot be followed much beyond the time frame of program delivery, and outcomes measured at young ages usually cannot be readily monetized. In addition, intervention effects linked to future costs are typically subject to down-weighting through discounting and fade-out.

At the same time, lasting effects may rely on delivering services to children during key developmental periods, and research has demonstrated the importance of early intervention (Cunha and Heckman, 2008; Dodge et al., 2008; Barnett, 2011). The challenge is how to value outcomes with complex processes involving multiple dynamic and interacting factors. For example, important new research is examining the mechanisms by which noncognitive factors and personality influence long-term success (Almlund et al., 2011). Ideally, this research could help explain how these factors collectively influence future adult outcomes and how they are best measured in economic evaluations. For example, noncognitive factors could be more important at older ages and more malleable, making them a better candidate for intervention with older children. Today, however, Jones indicated that the role of noncognitive skills on long-term success is not represented in economic evaluations of programs for children.

The Potential of Research

The field will be greatly helped by research that establishes the links between outcomes in program evaluations and future direct or indirect costs and benefits. This research should be based on robust methodology, include multiple studies, and involve causal associations, Jones said. Once those links are determined, then some consensus is likely to develop as to what measures best represent early skills. If certain domains, such as early aggression or social skills, are found to have a stronger association with economic outcomes controlling for other factors, these areas could be prioritized. Research also could factor in how much these traits may fluctuate over time, the likelihood that they may change as the chronological gap between measured skill and economic outcome is increased, the influence

of different contexts for understanding these associations, and the varying characteristics of different populations.

Variable associations may be represented in terms of likelihoods for later states to occur. For example, an improvement in an early mental health outcome may increase the likelihood for high school completion. Around this likelihood, the potential for variation in causal influence must be understood. In this context, ranges of estimates are good. Policy makers may not like ranges, but they need to be factored into the overall sensitivity analysis of the economic evaluation.

“The future is bright,” Jones concluded. Economic evaluations for family, child, and youth programs will only get better in the coming years. But standards and consistent methodologies are needed to compare across studies, Jones stated, and researchers need to fully consider the possible impacts of effective programs. Thus, some collective organization of determining and promoting appropriate methods and measures would help researchers in the future.

4

Issues to Consider in Benefit-Cost Analysis

Two sessions at the workshop focused on several of the prominent methodological issues that arise when evaluating preventive interventions. Those presentations are combined in this chapter to explore in greater depth some of the questions that will have to be addressed in developing standards for benefit-cost analyses. During the course of the discussion, a topic that received special attention from several presenters was the kinds of research designs that can generate valid evidence.

IDENTIFYING CAUSAL ESTIMATES BY RESEARCH DESIGN¹

Randomized controlled trials are not always possible to conduct when evaluating preventive interventions. In those cases, different experimental designs must be pursued for generating causal knowledge, including regression discontinuity analyses, interrupted time series designs, nonequivalent control group designs, and single case design. Thomas Cook explored the potential of several research designs that can be employed in benefit-cost analyses.

He began by looking at the within-study comparison method, which in the past has often produced results very close to those of randomized controlled trials. In one form, an overall population is selected into a randomized experiment group and a comparison group. The randomized experiment group is then randomly assigned to a control group and a treatment group, yielding a randomized controlled trial. The nonequivalent comparison groups are formed by systematic assignment into the same treatment

¹ This section summarizes information presented by Thomas Cook, Ph.D., Northwestern University, Evanston, Illinois.

and control statuses, with individuals most often self-selecting into the treatment or control status or being so appointed by an administrator. An effect size can be calculated for the observational study and compared with the effect size from the randomized experiment.

In the second form, a within-study comparison involves taking a randomized experiment, calculating its effect size and then contrasting a nonequivalent comparison group by self-selection or administrator selection, Cook continued. This nonequivalent comparison group is then linked to the same treatment groups as in the randomized experiment, and an adjusted effect size is created after trying to control for any selection differences between the treatment and nonequivalent comparison group. Then, the effect sizes from the experiment and nonexperiment are compared. If they are similar, the conclusion is drawn that the nonexperimental results were not biased. If they are different, the opposite conclusion is drawn.

Whatever within-study comparison method is chosen, an important goal is to identify the conditions under which a given observational method produces better or worse approximations to the causal estimates from a randomized controlled trial. For example, seven within-study comparisons comparing randomized controlled trials with regression discontinuity studies have produced similar causal estimates at the cutoff, Cook said. Though this is theoretically trivial, since regression discontinuity studies are supposed to produce this result, it is encouraging that the same cutoff-specific results were achieved as in the randomized controlled trials since this indicates that the implementation of the regression discontinuity studies did not produce bias. More important is the one relevant study comparing experimental results to those from a comparative regression discontinuity study that includes pretest values along all the assignment variable and so can serve to index the functional form relating the assignment variable to the outcome without treatment. This one study showed the same results as a randomized controlled trial in all areas away from the cutoff, and not just at it as in the simpler regression discontinuity design with only posttest data (Cook et al., 2008).

In addition, seven studies have compared randomized controlled trials with interrupted time series studies, all but one of which used a comparative interrupted time series with a nonequivalent control group time series. As Cook observed, all showed the same results for treatment-caused changes in the mean and, where tested, for changes in slope also.

Nonequivalent control group designs are more common than regression discontinuity or interrupted time series in actual research practice.

More than 20 within-study comparison studies of such designs exist in many different fields, though they are skewed toward job training and education reform. They show close approximation of randomized controlled trial results when the comparison groups are very local, when pretest measures of outcomes are used to match groups, when there is independent knowledge of all or most of the selection process, and when there is multilevel matching. It also works when there is a hybrid matching strategy that combines local matching where it works well with non-local focal matching wherever local matching does not work. These procedures can reduce all bias sometimes and some bias almost always, said Cook. However, none can be guaranteed to reduce all bias all the time. In the future, this line of research will continue to seek a combination of various strategies to improve predictions when randomized controlled trials results are stably, but not invariably, replicated.

Experimental Designs and Clearinghouses

Today, several clearinghouses exist, including the What Works Clearinghouse (WWC) (<http://ies.ed.gov/ncee/wwc>), the Campbell Collaboration (<http://www.campbellcollaboration.org>), and Blueprints for Healthy Youth Development (<http://www.blueprintsprograms.com>). Cook limited his remarks to the WWC, but other speakers, as described later in this chapter, spoke more broadly about the stamp-of-approval mechanisms applied by these organizations.

The WWC's standards for regression discontinuity studies are now being revised, Cook noted, but the currently proposed revisions do not deal with the advantages of comparative discontinuity studies, which "seems odd." However, steps are being taken to add consideration of such designs as the new standards are worked out. In addition, the WWC does not include interrupted time series or comparative interrupted time series except in single case designs, which seems "short-sighted given the evidence to date," according to Cook.

Today, the WWC accepts nonequivalent control group designs only if the treatment and control groups do not vary at pretest on measures of the outcome. But this is "naïve," said Cook, because it does not consider time-varying pretest mean differences by group and because some de-

signs without pretest measures of the outcome still use sophisticated matching techniques. More sophisticated consideration could be given to the WWC standards for accepting nonequivalent control groups as acceptable. However, Cook concluded by agreeing with almost all that is in the WWC, and he generally endorsed its standards. At issue for him are extensions to the current standards more than replacement. In particular, greater consideration could be given to the special status of comparative regression discontinuity studies and to comparative interrupted time series designs into the clearinghouse.

Results also would benefit by an external warrant, said Cook, so they are not the product of an individual researcher's choice or preferences. Current statistical theory cannot provide this warrant because it cannot specify those features of a given application that meet its assumptions. So when an experiment is not possible, "satisficing" standards could meet a threshold of acceptability. Within-study comparisons seek to generate an empirical warrant for meeting a "satisfying" criterion. Any imperfect design is satisfying if it often produces the same results as a randomized experiment on the same topic.

DESIGNING ERROR-TOLERANT STUDIES²

Sometimes policy makers use the results of benefit-cost analyses well, but in many cases research fails to translate into socially beneficial action, said Jens Ludwig. Part of the problem is the limited scientific literacy of the users of research. It is very difficult to explain to nonspecialists the difference between failing to reject the null hypothesis and accepting the null hypothesis or the difference between a good design and a bad design within a class of research. Furthermore, as Ludwig has learned from living in Chicago, "Politics turns out to be very political."

As an example of the difficulties encountered when disseminating the results of research, Ludwig pointed to the National Head Start Impact Study, which looked at the impact of Head Start on children at the end of first grade. Magazine reporters and think tank commentators used the results to conclude that Head Start does not work or delivers broken promises (Besharov, 2005; Barnett and Haskins, 2010; Klein, 2011). In

² This section summarizes information presented by Jens Ludwig, University of Chicago, Illinois.

fact, the results showed positive results in letter identification, spelling, vocabulary, and oral comprehension, but the confidence intervals were wide enough that two of the four results included the possibility of zero. Because the confidence intervals included both effects that were large enough for the Head Start analysis to pass a benefit-cost test, but also included zero, in the final analysis the impact study simply was not able to determine whether Head Start would pass a benefit-cost analysis. Ludwig said that this was a case of misunderstanding basic statistics.

One way to reduce misinterpretations of research results would be to educate the consumers of results. But if even think tank commentators misunderstand such results, said Ludwig, large-scale consumer education seems implausible.

Research Geared Toward the Research Consumer

Researchers need to anticipate the difficulty policy makers and consumers of research have in using research results in socially productive ways, Ludwig said. In fact, according to Ludwig it is very difficult for research consumers to distinguish between good and bad studies even within a given research design. Moreover, it is not all that clear that the existing “stamp of approval” mechanisms are helpful in guiding research consumers. One approach that the research field may consider in order to aid research consumers is to adopt an alternative system that simplifies the process of determining the difference between the good and bad studies by using a high-quality study design bar, which includes fewer classes of research designs. The trade-off for this approach is excluding some good studies that may fall short of the quality bar that is set for research designs. Some suggestions Ludwig made about what researchers can engage in to assist research consumers are mentioned below.

First, researchers should not conduct underpowered experiments, Ludwig said. In particular, their experiments should have adequate power to tell whether an intervention passes a benefit-cost test. Second, research should be reported in a way that the consumers of research are able to adjudicate between a good and bad study within a given research design class. One way for the research field to enable such judgments is through a stamp-of-approval mechanism. Today, several such mechanisms exist, Ludwig observed, including the WWC. At the moment, however, none of the existing mechanisms does exactly what is needed

to aid the consumer of research. These systems could be studied to determine whether they are constraining claims about evidence in ways that are helpful or unhelpful, though this determination is likely to differ from one policy area to another. An interesting question is whether all stamp-of-approval mechanisms could cohere on a given standard and whether that would solve the problem.

Ludwig stated that research consumers like having mixed results within a class of research designs. For instance, if policy makers or advocacy organizations know what they want the answer to be, they can look at a broad class of research studies to find the results that meet their expectations. Perhaps one way to solve this problem, said Ludwig, is to establish tight standards for research design quality.

One approach could be to include as evidence only the results of randomized controlled trials or regression discontinuity studies. A limitation to this approach, admitted Ludwig, is that such a decision would throw away good information from other kinds of studies. Another approach would be to develop a checklist that studies must satisfy. For public debates, the idea that some sort of checklist is going to constrain the kind of evidentiary claims that are made is probably not realistic.

Focusing on the errors made by the users of research results could yield important advances, Ludwig concluded. A useful step may be to do retrospective reports of how benefit-cost analyses are used, and abused, in the real world to identify problems that then could be mitigated. This could perhaps guide the field in how to move forward.

DECIDING WHAT EVIDENCE TO INCLUDE³

Rebecca Maynard discussed four key considerations surrounding what evidence to include in benefit-cost analyses: (1) overall relevance of the study, (2) relevance of the impact estimates that are reported, (3) causal validity of the impact estimates, and (4) adequacy of reporting.

³ This section summarizes information presented by Rebecca Maynard, Ph.D., University of Pennsylvania, Philadelphia, Pennsylvania.

The Overall Relevance of the Study

The first question is whether the intervention is relevant to the policy or practice decisions under consideration, said Maynard. Studies may address important interventions, but those studies may not be relevant in a particular setting or at a particular time. In a health intervention, for example, a set of medications or foodstuffs that need refrigeration have little or no relevance in areas where refrigeration is not available.

Another question regarding the overall relevance of a study is what an intervention is being compared with. For example, a social and emotional behavior program is going to have quite different effects in a functioning family than in a family in chaos. Knowing the counterfactual is important in judging whether or not a given set of impact estimates is relevant.

A third question involves the context. Was a study conducted in a place and time relevant to a conclusion? For example, Maynard argued that the Perry Preschool Study does not have great relevance today because the world is so different. Though it was one of the field's greatest impact evaluations, Ypsilanti, Michigan, in the mid-1960s was a very different place than today's world, where 70 percent of mothers are in the workforce and early childhood programs are much more common.

The Relevance of the Impact Estimates

Even if a study has relevance to a policy or practice problem, there may be issues that affect whether and, if so, how one might apply the findings. For example, does the reference period for the impact estimates match the intervention being considered, whether by a local school board, a state legislature, a parent group, or the federal government? The benefit-cost analysis could hone in on the benefits and costs that are of interest to a particular group.

A related issue is the relevance of particular estimates to the desired mission, Maynard pointed out. Does an estimate match the programmatic or policy goals? For example, someone in charge of child care policy in a state is probably not in a position to advance a program oriented toward crime prevention even if that program has good child care elements. Studies may also be relevant to a higher-level organization or may consider only the marginal treatment impacts rather than the overall impacts. It mat-

ters whether an impact estimate pertains to a primary or a secondary outcome of a study. For instance, results for secondary outcomes are more likely than those for primary outcomes to be selectively reported if the findings are favorable and ignored if the findings are null or unfavorable.

Exploratory analyses of data can inform theory and the next generation of research, but they are risky as a guide to the development of a logic model, Maynard observed. Some differences between treatments and controls will show statistical significance simply by chance. It is risky to include those measures in the same category as the outcome measures that were front and center in the original analysis.

The Causal Validity of the Impact Estimates and Reporting

Even studies that start out as well-executed randomized controlled trials can suffer from attrition (including differential attrition), measurement shortfalls, and analysis and reporting shortfalls. Matched comparison groups may differ because of untestable assumptions. Even with randomized controlled trials, the method of creating the intervention and control groups may be flawed. Finally, are the analysis methods appropriate to the sample design? Do they adequately address selection issues? “You will be surprised what happens,” said Maynard. “Really smart people sometimes do really silly things.”

Outcome measures should be based on meaningful metrics like dollars or percentile rankings, Maynard emphasized. They also need contextual information, such as the characteristics of the study sample, the intervention characteristics, and intervention context, Maynard said, along with information about the implementation of the intervention to know the fidelity with which it was carried out.

Maynard’s final plea was to eliminate standardized mean differences from reports of impact estimates. An effect size is a unit of currency, she said. An effect size that uses the standard deviation for a national sample of children is not in the same units as an effect size that uses all of the children in the bottom income quartile or who are English language learners.

Clearinghouses for the Results of Benefit-Cost Analyses

One function of clearinghouses is to reflect standards as to the kinds of studies and results that are judged to have credible evidence on particular topics. In that respect, said Maynard, it is important and useful to build on the standards being applied by existing clearinghouses. For example, the review process for the WWC includes development of a review protocol to ensure alignment and consistency, identification of the relevant literature to promote consistency and completeness, screening and reviewing of the studies using a consistent format, summarizing the findings, and archiving the findings in a shareable format. Sharing coded data and studies is particularly important, said Maynard, so that information is available to others. The WWC has been in operation for 10 years and is gaining acceptance. Researchers are designing their studies to the higher end of the standards to avoid reservations about the credibility of the reported findings. Studies are being reported out better, and data-sharing agreements are facilitating the reuse of data. It may be interesting to adopt something like the WWC standards as a “base” on which other agencies could build their evidence databases. Other agencies or those who use the standards could then tailor the standards for their own use and dissemination.

Maynard also pointed out that an interagency workgroup in the federal government is now thinking about whether there should be a common evidence platform where coded studies could be housed and information made available. Standards differ somewhat from one organization to the other. Should they be reconciled? How should their characteristics be formatted? Every standards-setting organization does not need to be the same, but where they differ should be known so differences could be mediated if necessary.

ISSUES WITH RANDOMIZED CONTROLLED TRIAL DESIGNS

Several of the speakers discussed randomized controlled trials during the question-and-answer sessions. Basu expressed his concern about calling randomized controlled trials the “gold standard” for benefit-cost analysis. Such trials can have many problems, he noted, both in their design and in the data they produce. Perhaps one approach would be to

identify a checklist of criteria that alternative study designs would need to meet to be accepted.

Cook agreed that randomized controlled trials are not a gold standard if a gold standard implies being infallible. They have their own assumptions, and even a randomized clinical trial does not guarantee internal validity. Bias can arise from small selection differences, from small violations of the separate condition, and from other factors. Cook asserted that they are better than the alternatives, at least in terms of internal validity. All the other alternatives, except for regression discontinuity require full knowledge of the selection process.

Maynard agreed that the term “gold standard” is inflammatory and should not be used. Randomized controlled trials are not necessarily perfect or even better than some alternative designs, she said, because they require much more than just effective randomization. The real question is how good is the evidence. Standards could help determine whether a study is relevant or not relevant to the question at hand.

INCREASING THE COMPARABILITY OF BENEFIT-COST ANALYSES⁴

The ultimate goal of benefit-cost analysis is not just to look at a given program and decide whether it has a favorable economic return, said Lynn Karoly; it is to compare programs across sectors. In the early childhood sector, for example, policy makers have choices among different early childhood programs, but they also are making choices among early childhood programs, school-based programs, and prevention programs versus remediation programs. In this way, they can develop a portfolio of investments that can yield an optimal investment strategy. One of the reasons for developing standards for benefit-cost analyses is to enable such choices.

The lack of standards in benefit-cost analyses can result in the creation of differing messages. In the area of early childhood education, for example, benefit-cost analyses for six early childhood programs show a broad range of returns (see Figure 4-1). Even for the same program

⁴ This section summarizes information presented by Lynn Karoly, Ph.D., RAND Corporation, Arlington, Virginia.

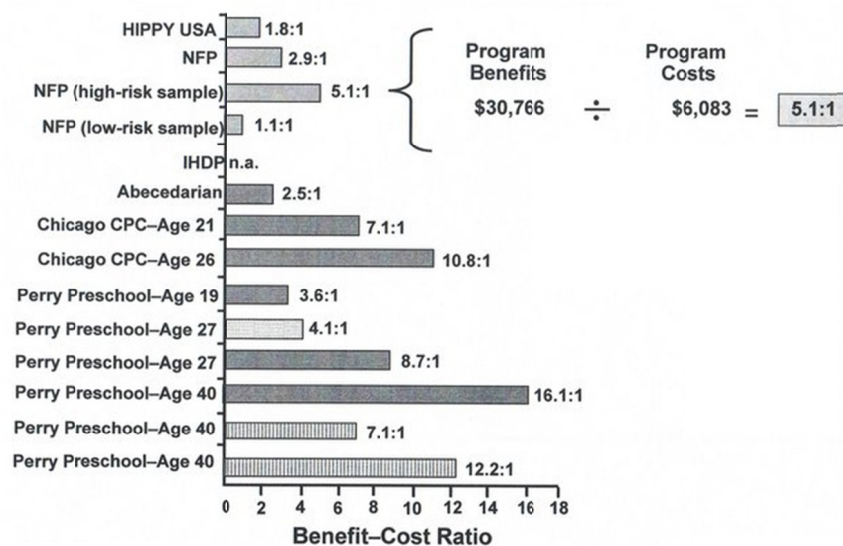


FIGURE 4-1 The reported benefit-cost analysis ratios for early childhood interventions vary substantially.

NOTE: CPC = Child-Parent Center; HIPPY = Home Instruction Program for Preschool Youngsters; IHDP = Infant Health and Development Program; NFP = Nurse-Family Partnership.

SOURCE: Adapted from Karoly, 2012.

following subjects to the same age, results can vary markedly when different methods are used.

A basic question is how much of the differences derive from choices the analysts are making in performing benefit-cost analysis, Karoly observed. The studies differ along a number of dimensions, including the age at follow-up, discount rate, discount age, whether the results are disaggregated by stakeholder, and whether standard errors are reported (see Table 4-1). For example, some of the studies discount the benefits and costs to age three, while others discount them to birth. The studies also use a wide range of discount rates. Further, studies do not always disaggregate the results by stakeholders, such as the government, taxpayers, or society as a whole.

All of the studies have performed some kind of sensitivity analysis, Karoly continued. For example, they may have disaggregated estimates of benefits to show the different sources of benefits, such as the benefits

TABLE 4-1 Variation in Methods in Application of Benefit-Cost Analysis to Early Childhood Programs

Program	Age at FU	Discount Rate(s)	Discount Age	Disaggregate by Stakeholder	SEs	Sensitivity Analysis
HIPPY	6	3	3	Yes	Yes	Some
NFP (WSIPP)	15	3	0	Yes	Yes	Some
NFP (RAND)	15	4, 0–8	0	Yes	Yes	Some
IHDP	8	3	0	Yes	Yes	Some
Abecedarian	21	0, 3, 5, 7, 10	0	No	No	Some
Chicago CPC	21	3, 0–7	3	Yes	No	Some
Chicago CPC	26	3, 0–7	3	Yes	Yes	Some
Perry–Age 19	19	3	3	Partial	Yes	Some
Perry–Age 27	27	3, 5, 6, 9, 10, 11	3	Partial	No	Some
Perry–Age 27 (RAND)	27	4, 0–8	3	Yes	Yes	Some
Perry–Age 40	40	0, 3, 7	3	Partial	No	Some
Perry–Age 40 (Heckman)	40	0, 3, 5, 7	3	Yes	Yes	Some

NOTE: CPC = Child–Parent Center; FU = follow-up; HIPPY = Home Instruction Program for Preschool Youngsters; NFP = Nurse–Family Partnership; SEs = standard errors.

SOURCE: Adapted from Karoly, 2012.

to be gained by a reduction in crime. Decision makers then can adjust their use of the results based on the sensitivity analysis. In other cases, sensitivity analyses look at the effects of basic assumptions in the model, such as the size of the dead weight loss from taxation. Similarly, most of the studies report standard errors, though some also report the percentage of time that a positive net benefit would be expected.

Establishing Base Cases

One way to increase the comparability of benefit-cost analyses, both within the early childhood domain and across policy areas, would be to identify a base case or standard case on which analysts could agree, Karoly said. Researchers then could make the case for an alternative analysis of the data but only after presenting results for the base case. For example,

regarding the appropriate real discount rate, a base case can be used along with a sensitivity analysis using alternative discount rates. A similar approach could apply to the preferred age to discount costs and benefits. With regard to uncertainties, in addition to reporting standard errors, the percentage of simulations with positive net benefits could be reported in the base case, a useful and intuitive result for decision makers. Other areas of uncertainty, such as the methods for projecting future benefits, assumptions about the efficacy of scale-up, or distributional weights, could be better addressed with sensitivity analyses.

Karoly also urged that attention be given to the proper outcomes or summary measures from benefit-cost studies. If only a benefit-cost ratio is reported, it can be difficult to recover the numerators and denominators. In addition, measures like the internal rate of return or benefit-cost ratios will not necessarily order projects in the same way as would net benefits. These issues could be addressed by the use of a reliable internal rate of return or by adjusting the benefit-cost ratio for projects of different size.

Karoly emphasized that the analysis of costs produces value in its own right, even if it does not lead to a benefit-cost analysis. If a good cost analysis were done as part of every program implementation, it would be available should a benefit-cost analysis be warranted. She also emphasized the role that administrative data can play, both in short-term evaluations and in learning about long-term impacts. If assessment of potential long-term impacts is built into a program evaluation from the beginning by linking to administrative records, evaluation costs would be lower.

Standardization may be easier in some areas than in others, Karoly acknowledged. The easier issues to address include valuing program costs, the discount rate, the age to discount to, accounting for uncertainty, sensitivity analysis, disaggregation by stakeholder, and reporting results. Harder issues to address include the right baseline to use, the length of follow-up, the outcomes measured, the use of shadow prices, and the projection of future outcomes for participants, family members, peers, and descendants.

EXPRESSING UNCERTAINTY IN BENEFIT-COST ANALYSES⁵

As Anirban Basu pointed out, expressions of uncertainty are rare in policy analysis. In particular, both cost-effectiveness and benefit-cost analyses often report their results as point estimates without expressions of uncertainty. Yet policy predictions often rest on unsupported assumptions or leaps of logic, rendering expressions of certitude not credible (Manski, 2011).

Though no empirical evidence exists on the issue, researchers tend to assume that policy makers do not want to know about uncertainties. As Lyndon Johnson once said, “Ranges are for cattle. Give me a number.” But even if a decision maker does not want to know about uncertainty, the uncertainty remains important, Basu insisted, because the production function for the data is often nonlinear with respect to its many inputs. With a nonlinear function, the expectation of the function is not equal to the function of the expectation of inputs. Thus, using point estimates as the input to a model can generate a very different result than when incorporating the uncertainties around those inputs.

Forms of Uncertainty

Another way to look at a production function is to recognize that it is inherently heterogeneous, said Basu. This heterogeneity can be divided into scientifically known heterogeneity, such as age effects or gender differences, and scientifically unknown heterogeneity (the unknown unknowns), which is known to exist but cannot be explained. The scientifically known heterogeneity can in turn be divided into heterogeneity that is directly observable for the decision maker or analyst (the known knowns) and heterogeneity that cannot be observed (the unknown knowns).

The unknown unknowns and unknown knowns together create stochastic uncertainty, parameter uncertainty, and structural uncertainty in a model, Basu observed. Stochastic uncertainty is used to study random variation across individual outcomes, often using Monte Carlo techniques. Parameter uncertainty is used to study variation in expected outcomes through probabilistic approaches, especially second-order Monte

⁵ This section summarizes information presented by Anirban Basu, Ph.D., University of Washington, Seattle.

Carlo simulations. Finally, structural or model uncertainty is used to study variation in expected outcomes and individual outcomes. There is no clear consensus on how to study such uncertainty, though common approaches are sensitivity analysis and model averaging.

Unexplained variations in expected outcomes lead to decision uncertainty, said Basu. Decision uncertainty can be expressed in different ways, depending in part on what a decision maker wants to know. One approach is to give 95 percent confidence intervals for a benefit-cost ratio. Another is to give a 95 percent confidence interval for net monetary benefits. A third is to provide acceptability curves, which relate the probability of acceptance of an intervention to the willingness to pay for a given benefit at a given cost. Each of these three methods provides exactly the same information about uncertainty in benefit-cost or cost-effectiveness analyses. Which is chosen depends on the research and on the decision maker's comfort in understanding heterogeneity.

Acceptability curves can be linked directly to decision making for future research, Basu pointed out. The expected value of perfect information about a policy is the product of a probability that a decision made today is wrong multiplied by the loss due to a wrong decision. This simple product provides an upper bound on the value of future research and can be used to inform many decisions, such as whether to fund future research, how to prioritize across future research proposals, and how to design future research studies.

THE POTENTIAL COMPENSATION TEST AND DISCOUNT RATES⁶

Richard Zerbe discussed two contentious topics in the area of benefit-cost analysis. The first is the idea of the potential compensation test and the second is on discount rates. The first holds that if the winners from a project could in theory compensate the losers, without considering interpersonal comparisons of utility such as income distribution, then the project is beneficial. This criterion is not satisfactory, said Zerbe, for several reasons.

⁶ This section summarizes information presented by Richard Zerbe, Ph.D., University of Washington, Seattle.

First, the idea that such compensation could exist is clearly a fiction, Zerbe pointed out. In fact, the cost of compensating losers in a particular project would in many if not most cases be greater than the net value of the project itself. Second, Zerbe explained that the potential compensation test does not always work in the realm of law. When disputes arise, the sum of the expectations of the parties generally exceeds the total value available. A potential compensation test therefore would not succeed. Nevertheless, benefit-cost principles can be used to make decisions in that case without becoming hung up on the potential compensation test. Third, the potential compensation test leads to what are known as Scitovsky reversals, in which moving from state A to state B is beneficial, but moving from state B to state A also is beneficial.

Finally, a better alternative is available, Zerbe stated. When considering a portfolio of benefit-cost projects, a person who loses and pays taxes to support one project can gain from another project. Because each project can be expected with a probability greater than 50 percent to have net gains, in the end almost everyone wins. This is true even if some projects do not pass a benefit-cost test or if some gaming of the system occurs.

Discount Rates

The second topic Zerbe discussed is discount rates. A wide range of discount rates can be justified, from 0 to 30 percent or more, according to Zerbe. Furthermore, the literature on discount rates is confusing and disparate.

One rate that has been suggested is the rate of return to private investment, which Zerbe and his colleagues have calculated to be about 8.5 percent in real terms. Other approaches are to use a weighted average of the rate of return and time preference rates, a social welfare function, or a time-declining rate, which can be combined with the other approaches. Investment in a public project displaces some private capital, and some of the investment comes from a reduction in consumption. The proportions that come from private capital and the proportions that come from a reduction in consumption can be used to calculate a rate that represents the opportunity cost of capital, which Zerbe has calculated to vary between 6 percent and 8 percent.

Zerbe added that the consumption rate of interest, which can be thought of as a pretax private return to individuals, is generally calculated to be about 3.5 percent. It is too expensive to anticipate for each project where the funding is coming from exactly, so 6 to 8 percent is a project average. Zerbe indicated that a case can be made for rates as low as 3.5 or as high as 8.5 for particular projects.

The only way to develop a rough consensus on discounting, Zerbe concluded, is to develop fundamental principles that can be used to decide on the rate. One such principle is that ethical considerations and other extraneous considerations, such as environmental goods or the value of life, should be excluded from the discount rate. These factors can be included in values and thus in benefit-cost analyses through willingness to pay tests, but they should not be included in discount rates. A second principle is that a discount rate should require that no project be accepted if its return is less than the return available on alternative projects. This is a straightforward opportunity cost rationale, Zerbe said.

5

Translating Results to Inform Policy and Practice

The objective of benefit-cost analysis is to bring evidence to bear on the policy making process. Four speakers at the workshop explored that objective from the perspective of the users of benefit-cost analyses. All agreed that benefit-cost analyses have provided valuable guidance to policy makers.

A PERSPECTIVE FROM THE OFFICE OF MANAGEMENT AND BUDGET¹

For decades the U.S. Office of Management and Budget (OMB) has sought to make the best use of the marginal dollar of federal funding, said Kathy Stack. Yet in only a few areas does evidence exist for how to do that. According to Stack, government programs are marked by immense inertia. Federal agencies are quite comfortable doing what they did the year before, unless somebody tells them to do something differently. Also, Congress places a priority on maintaining political support. As a result, taking money away from any particular entity is very difficult unless a compelling reason exists for doing so. Finally, organizational silos cause communications to break down, creating missed opportunities where organizations are not talking to each other about ways of doing things better.

The budgeting process is also characterized by what Stack called the “wrong pockets” problem. For example, interventions in housing could produce savings in health care, but appropriations subcommittees are focused on housing and on health care, not on the connections between the two. Similarly, appropriators for discretionary programs in govern-

¹ This section summarizes information presented by Kathy Stack, U.S. Office of Management and Budget, Washington, District of Columbia.

ment are unlikely to change existing priorities to produce savings in mandatory programs. The same thing happens between the federal government and the states, which often operate at cross-purposes rather than collaboratively.

A Brief History of Evaluation in the Federal Government

A major challenge to breaking out of entrenched practices has been the lack of robust measurement and evaluation tools, Stack observed. In the 1980s the effect of a program focused on children was measured simply by the number of children served. In the 1990s the Government Performance and Results Act shifted the emphasis to inputs, outputs, and outcomes. Operationalizing those concepts within agencies turned out to be exceedingly difficult. In seeking to do comprehensive assessments of all federal programs, the Bush administration developed the Program Assessment Rating Tool, which required agencies to evaluate every program, largely on the basis of performance data. This generated a lot of work on defining outcomes and outputs, Stack reported, but the data produced often did not say much about impacts or cost effectiveness.

At the beginning of the Obama administration, stimulus money became available to think about new program designs, which led leaders in the Executive Office of the President to look for opportunities to embed research in new program designs, according to Stack. For example, a presentation by David Olds about the Nurse–Family Partnership program to a group of staffers at the OMB during the Bush administration contributed eventually to the development of a \$1.5 billion program by the Obama administration. Similarly, the OMB has emphasized ways of building evidence through comparative cost effectiveness in its communications with agencies.

Successive administrations also have sought to build a clearinghouse to create a repository of knowledge about impacts and cost effectiveness. The WWC is one product of this emphasis, and several others have been developed. However, the clearinghouses lack some forms of data, such as cost data and comparative data. Standards that could serve as a “north star” for the convergence of agency discussions and actions would be helpful, Stack said.

A Waiting Audience

Benefit-cost analysis researchers have a “waiting audience” for their work, said Stack. Both the OMB and the budget committees on Capitol Hill have policy levers they can use to drive funding to programs that are more cost effective, but mustering the political will to change requires evidence. Federal policy makers tend to do things in 5- and 10-minute chunks, Stack explained. To have an effect, information needs to be simple and clear, even with complex research. A simple presentation that grabs people’s attention and focuses on outcomes can be influential.

In response to a question, Stack also pointed out that the OMB is eager to enable researchers to conduct evaluations of programs, whether through random assignment, quasi-experiments, the use of administrative data, or some other methodology. Some agencies are doing this well, while others are lagging behind. The idea is to seek out variation within programs and figure out which variants make the most difference. Though most policy makers are not focused on doing random assignment studies, virtually every large program has opportunities for doing such studies if program managers can be connected with researchers at the appropriate time.

The OMB is encouraging programs to generate data about effectiveness, Stack concluded, but evidence standards in addition to incentives could work to encourage people to adhere to those standards. If programs had strong evaluation components, they could learn as they go. Data on cost effectiveness and return on investment could redirect money at the state and local levels, indicated Stack.

A STATE POLICY PERSPECTIVE²

The term *policy maker* is usually interpreted to include legislators who make laws and fund programs, but it actually includes a wider range of people, said Jacqueline Jones, former Deputy Assistant Secretary for Policy in Early Learning at the Department of Education. Among these people are the program specialists who write regulations, implement programs, and monitor progress within the executive branch of government.

² This section summarizes information presented by Jacqueline Jones, Ph.D., independent consultant, Princeton, New Jersey.

The advocates and stakeholders who support or oppose legislation and regulation and aim to influence policy are also included. These policy influencers can be a powerful force on legislators and program specialists, said Jones.

Benefit-Cost Analyses in Practice

Before going to work at the U.S. Department of Education, Jones was assistant commissioner for the Division of Early Childhood Education in the state of New Jersey. In 2006 and 2007, New Jersey was implementing a court-ordered high-quality preschool program charged with providing all 3- and 4-year-olds in 31 of the state's poorest districts with a full-day full-year preschool program. The governor of the state and commissioner of education were very interested in expanding the program beyond the 31 districts to the more than 600 districts in the state and in changing the way the program was funded.

The preschool program had a number of components, including its full-day duration, a maximum of 15 children per class, a requirement that teachers have a bachelor's degree and a P-3 certification, the use of master teachers who have at least 5 to 7 years of experience, assistance from family workers, and transportation. Cost was a major consideration, which required considering the benefits from several of the program's components. For example, do data exist that 15 children in a preschool classroom is better than 17? Do the teachers need to have bachelor's degrees right away or can they earn their bachelor's degrees while they are teaching? Which parts of the program should be implemented first and which later to ensure high quality? Meanwhile, legislators did not know how they were going to pay for the program, the governor was determined to make it happen, and policy influencers were threatening court action to see that certain components were implemented.

As an example of the inevitable complications, Jones pointed out that teacher salaries were the biggest driver of cost, and the program was paying teachers at parity with public school teachers. But each district in New Jersey negotiates on its own with its union, so each district had a different pay scale. New Jersey "is a very complicated place," Jones said.

Interventions involve multiple actions and have multiple outcomes. A major question is therefore to what extent do components of a complex preschool program function independently and to what extent are they

interrelated in ways that are not yet understood. For example, would removing the master teachers have a major adverse effect on the program? What is the role of teacher preparation in the program? What contribution do family workers make to preschool education, and does that contribution differ from place to place?

At times, policy makers knew that the promises they were making were not entirely borne out by research. But pressures can be so intense that people will do whatever they can to make something happen, Jones said. Everyone wants better outcomes for children, but how to pay for programs and see results within the time frame of a particular administration is not easy.

Jones called for more conversation between policy makers and researchers. Policy makers usually need information right away, not a week from now. Researchers can benefit by knowing what policy makers need and what kinds of stresses they are under. For their part, advocates, in their zest to do the right thing, sometimes promise more than can actually be delivered, Jones observed. Social scientists could help policy makers and advocates understand what works for whom under what circumstances.

Suggestions for Action

Jones made several suggestions for increasing the value of benefit-cost analyses to policy makers. Clear and accurate description of what is happening with control groups is critical, she said. Also, gathering baseline data before starting to look for effects can improve the quality and usefulness of data. Then, standardizing the presentation of data can be tremendously important by providing a common conceptual framework with which people can interpret results. The world has changed since some of the landmark studies in the field were conducted. New studies are needed that reflect modern circumstances, Jones said.

Legislators can expect variations in program quality, especially in the early stages of implementation. No program is monolithic, even those that have sets of standards. Moreover, benefits do not appear immediately after a program is instituted.

Finally, researchers could ask policy makers what they think would be helpful. Ongoing conversations can help policy makers learn about the complexity of an issue, encouraging them to ask more informed ques-

tions, while helping researchers understand the issues that policy makers face. Such relationship building can foster a sense of trust and support on both sides.

A PERSPECTIVE FROM THE DEPARTMENT OF HEALTH AND HUMAN SERVICES³

Good policy results from a combination of three inputs, said Linda Smith. The first is good research about what works and what does not work. The second is good data. The third is human stories. Policy makers may know the research and have the data, but they may not take action without knowing about the human aspects of a program.

The Complexity of Using Research to Inform Policy

Benefit-cost analyses tend to focus on particular aspects of a program and ignore other aspects, Smith said. For example, studies of the Perry Preschool Project have focused on its effects on children. But an important aspect of that program was the parent involvement fostered through its weekly home visits, which has not received as much attention. These home visits changed the parent-child relationship, but it also changed the parents and their relationship with their communities. For example, these interventions can have an effect on parent's work and their ability to get out of poverty. These kinds of effects require a different lens to detect, said Smith, whether examining data from past studies or planning future studies.

Child development is extremely complex and cuts across social, emotional, cognitive, and other domains, Smith stated. Programs need both horizontal and vertical alignment to be maximally effective. No matter where a child is, according to Smith that child needs and deserves consistent interventions and a certain level of quality of care.

³ This section summarizes information presented by Linda Smith, Deputy Assistant Secretary and Inter-Departmental Liaison for Early Childhood Development for the Administration for Children and Families at the U.S. Department of Health and Human Services, Washington, District of Columbia.

Research results also can be misinterpreted, Smith observed. Researchers may assume that their results have clear implications. Unfortunately, this assumption is not always warranted. Researchers can help policy makers not to use research to make bad decisions.

Making Policy Decisions with Limited Budgets

Benefit-cost analyses can be hugely important in implementing and sustaining a program, and more are needed, particularly as decisions are made about the future of Head Start and child care programs in the United States, stated Smith. Given that policy makers need to make decisions with the information available and limited resources, these can be very difficult decisions to make. What is happening with child care, for example, is that quality is declining as funding fails to keep up with the need, because no one is willing to cut the number of child care slots, Smith said.

The reauthorization of Head Start in the next few years will entail making these kinds of hard decisions. For example, mandating a longer day, a longer week, or a longer year has enormous budget implications. Which of these three options would return the most benefits? Today, teachers in Head Start make approximately half as much as teachers in the public school system. A tough decision that has to be made is whether to increase the pay of teachers or cut the number of children served. The results of sequestration have been severe in Head Start, Smith reported, and program managers have trimmed as much as they can, leaving wages not much above the poverty line.

Smith indicated that another decision involves how much of a program an individual child should receive. Is it better for 1 child to get 2, 3, or 4 years of child care, or is it better for more children to get 1 year? Similarly, are programs targeted to poor or mostly minority children preferable to programs that include children from more advantaged backgrounds? These are the kinds of decisions that managers are now facing.

In almost all of these areas, program managers do not have much data with which to make decisions. For example, Head Start has retained its emphasis on parent engagement, but what about parent engagement is most important? Could a different form of engagement have greater benefits?

Communicating Research Results

For research to influence policy, it needs to be understandable, Smith emphasized. Whenever research is not translated into simple language, an opportunity is lost. Policy makers and policy implementers struggle with interpreting the results of benefit-cost analyses for the general public. Yet, without public backing, better policies are hard to implement.

A PERSPECTIVE FROM THE STATES⁴

Gary VanLandingham, director of the Pew-MacArthur Results First Initiative, which is a joint program of the Pew Charitable Trusts and the John D. and Catherine T. MacArthur Foundation, said that the initiative is essentially trying to replicate a model developed by the Washington State Institute for Public Policy. The initiative is working with 14 states and 2 counties in California to replace Washington State-specific data with data specific to other locations to provide benefit-cost analyses to policy makers in a form that they can use.

The Results First Initiative also has conducted a nationwide assessment of the field, focusing at the state level. It recently issued a report based on a comprehensive assessment of cost-benefit analyses produced by the 50 states and the District of Columbia during a 4-year period. This assessment identified about 1,000 studies that looked like cost-benefit analyses and then closely analyzed a third of those. It looked for six features:

1. Did they measure program cost and benefits across some type of baseline?
2. Did they assess both direct and indirect costs?
3. Did they discount future costs and benefits to current year values?
4. Did they monetize tangible and intangible benefits?
5. Did they disclose assumptions?
6. Did they do some form of sensitivity analysis?

⁴ This section summarizes information presented by Gary VanLandingham, director of the Pew-MacArthur Results First Initiative, a joint program of the Pew Charitable Trusts and the John D. and Catherine T. MacArthur Foundation, Washington, District of Columbia.

Of the 384 studies assessed, only 11 percent met all 6 criteria, leading VanLandingham to conclude that substantial room for improvement exists. Full monetization of tangible and intangible benefits was by far the weakest criterion.

Finding a Baseline of Practice to Inform Policy

Policy makers cannot wait several years for research results to be available, according to VanLandingham. They need to make decisions in real time. VanLandingham indicated that clearinghouses therefore can play a critical role by collecting and disseminating information being produced by cost-benefit analyses. However, a challenge of using clearinghouses is the differing nomenclature they use. Of eight clearinghouses reviewed by the Results First Initiative, the best tier of programs are alternately called *well supported*, *top tier*, *effective*, *proven*, *positive*, *strongly positive*, or are given a score of 3 to 4. Agreeing on what to call the good programs would be a step forward, said VanLandingham.

Policy makers have a great hunger for this kind of information, observed VanLandingham. In addition, advocates have latched onto it as a way to promote favored programs. They often cite studies that demonstrate returns on investments as a way to influence policy.

However, that influence has a flip side, VanLandingham warned. The credibility of the field could be destroyed unless a baseline of practice is established. Already, there is little relationship between the technical quality of benefit-cost analyses and their use in the policy process. Policy makers are often using all studies alike regardless of how they are viewed within the field. Even if consensus in some areas is difficult to achieve, agreement on several fairly basic things would help maintain the field's credibility by establishing standards for practice.

VanLandingham pointed out that other organizations engage in this kind of standard setting. For example, the Government Accounting Standards Board establishes standards for accounting to help people agree on the validity of an accounting statement. Standards setting does not happen overnight, but if the field can get started and move through an iterative process, it can maximize its impact on policy.

References

- Almlund, M., A. L. Duckworth, J. J. Heckman, and T. D. Kautz. 2011. Personality psychology and economics. In *Handbook of the economics of education*. Vol. 4B, edited by E. A. Hanushek, S. Machin, and L. Woessmann. Amsterdam: Elsevier. Pp. 1–181.
- Aos, S. 2013. *Benefit-cost information for public policies, the Washington State legislative approach: Some principles learned in the policy trenches*. Presentation at IOM/NRC Workshop on Standards for Benefit-Cost Analysis of Preventive Interventions for Children, Youth, and Families, Washington, DC. <http://iom.edu/~media/Files/Activity%20Files/Children/AnalysisofPreventiveIntervention/Panel%203%20Aos.pdf> (accessed February 14, 2014).
- Aos, S., M. Miller, and E. Drake. 2006. *Evidence-based public policy options to reduce future prison construction, criminal justice costs, and crime rates*. Olympia, WA: Washington State Institute for Public Policy.
- Barnett, W. S. 1985. Benefit-cost analysis of the Perry Preschool Program and its policy implications. *Educational Evaluation and Policy Analysis* 7(4):333–342.
- Barnett, W. S. 2011. Effectiveness of early educational intervention. *Science* 333:975–978.
- Barnett, W. S., and R. Haskins. 2010. *Finally, the Obama administration is putting Head Start to the test*. Washington, DC: Brookings Institution.
- Barnett, W. S., and L. N. Masse. 2007. Comparative benefit-cost analysis of the Abecedarian program and its policy implications. *Economics of Education Review* 26:113–125.
- Belfield, C. R., M. Nores, S. Barnett, and L. Schweinhart. 2006. The High/Scope Perry Preschool Program cost-benefit analysis using data from the age-40 follow-up. *Journal of Human Resources* 41(1):162–190.
- Besharov, D. J. 2005. *Head Start's broken promise*. Washington, DC: American Enterprise Institute.

- Cook, T. D., W. R. Shadish, and V. C. Wong. 2008. Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management* 27(4):724–750.
- Crowley, D. M., D. E. Jones, M. T. Greenberg, M. E. Feinberg, and R. Spoth. 2012. Resource consumption of a diffusion model for prevention programs. *Journal of Adolescent Health* 50:256–263.
- Cunha, F., and J. J. Heckman. 2008. Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources* 43(4):738–782.
- Dodge, K. A., M. T. Greenberg, and P. S. Malone. 2008. Testing an idealized dynamic cascade model of the development of serious violence in adolescence. *Child Development* 79(6):1907–1927.
- Drake, E. 2010. *Washington State juvenile court funding: Applying research in a public policy setting* (Document No. 10-12-1201). Olympia, WA: Washington State Institute for Public Policy.
- Durlak, J. A., and E. P. DuPre. 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* 41:327–350.
- Elliott, D. S., and S. Mihalic. 2004. Issues in disseminating and replicating effective prevention programs. *Prevention Science* 5:47–53.
- Feinberg, M. E., D. E. Bontempo, and M. T. Greenberg. 2008. Predictors and level of sustainability of community prevention coalitions. *American Journal of Preventive Medicine* 34(6):495–501.
- Jacobsen, R. H. 2013. *Hands-on guide to cost-benefit-analysis of crime prevention efforts*. Copenhagen, Denmark: Centre for Economic and Business Research.
- Karoly, L. A. 2012. Toward standardization of benefit-cost analysis of early childhood interventions. *Journal of Benefit-Cost Analysis* 3(1):1–43.
- Klein, J. 2011. Time to ax public programs that don't yield results. *Time Magazine*, July 7, 2011, <http://content.time.com/time/nation/article/0,8599,2081778,00.html> (accessed February 3, 2014).

REFERENCES

65

- Kuklinski, M. 2013. *The need for standards: Lessons from benefit-cost analysis of communities that care*. Presentation at IOM/NRC Workshop on Standards for Benefit-Cost Analysis of Preventive Interventions for Children, Youth, and Families, Washington, DC. <http://iom.edu/~media/Files/Activity%20Files/Children/AnalysisofPreventiveIntervention/Panel%201%20Kuklinski.pdf> (accessed February 14, 2014).
- Lee, S., S. Aos, E. Drake, A. Pennucci, M. Miller, and L. Anderson. 2012. *Return on investment: Evidence-based options to improve statewide outcomes* (Document No. 12-04-1201). Olympia, WA: Washington State Institute for Public Policy.
- Lutenbacher, M., W. O. Cooper, and K. Faccia. 2002. Planning youth violence prevention efforts: Decision-making across community sectors. *Journal of Adolescent Health* 30(5):346–354.
- Manski, C. F. 2011. Policy analysis with incredible certitude. *Economic Journal* 121(554):F261–F289.
- Miller, T., and D. Hendrie. 2008. *Substance abuse prevention dollars and cents: A cost-benefit analysis*, DHHS Pub. No. (SMA) 07-4298. Rockville, MD: Center for Substance Abuse Prevention, Substance Abuse and Mental Health Services Administration.
- NRC (National Research Council) and IOM (Institute of Medicine). 2009. *Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities*. Washington, DC: The National Academies Press.
- Pew-MacArthur Results First Initiative. 2012. *Better results, lower costs: Washington State's cutting-edge policy analysis model*. http://www.pewstates.org/uploadedFiles/PCS_Assets/2012/ResultsFirst_Washington_casestudy.pdf (accessed February 3, 2014).
- Reynolds, A. J., J. A. Temple, B. A. B. White, S. Ou, and D. L. Robertson. 2011. Age 26 cost-benefit analysis of the Child-Parent Center Early Education Program. *Child Development* 82(1):379–404.
- Zerbe, R. O., R. D. Plotnick, R. C. Kessler, P. J. Pecora, E. Hiripi, K. O'Brien, J. Williams, D. English, and J. White. 2009. Benefits and costs of intensive foster care services: The Casey Family Programs compared to state services. *Contemporary Economic Policy* 27(3): 308–320.

Appendix A

Glossary

attrition—in the context of research studies, refers to the gradual loss of study participants, some percentage of whom often drop out.

benefit-cost analysis—a method of economic analysis in which both costs and outcomes of an intervention are valued in monetary terms, permitting a direct comparison of the benefits produced by the intervention with its costs (also referred to as cost-benefit analysis).

cost-effectiveness analysis—a method of economic analysis in which outcomes of an intervention are measured in nonmonetary terms. The outcomes and costs are compared with both the outcomes (using the same outcome measures) and the costs for competing interventions, or with an established standard, to determine if the outcomes are achieved at reasonable monetary cost.

discount rate—a factor used to estimate future costs or the value of future benefits at the current equivalent value, used with the goal of attempting to take into account likely changes in valuation, opportunity costs, and other factors.

effect size—the magnitude of results (or effects on participants) of a particular treatment or intervention that is being studied.

interrupted time series study—is a type of quasi-experiment in which measures on a sample or a series of samples from the same population are obtained several times before and after a manipulated event or a naturally occurring event.

Monte Carlo simulation—the repeated drawing of uncertain parameters from assumed distributions to produce a distribution of possible outcomes. In benefit-cost analysis, Monte Carlo simulation is used to translate uncertainty in predicted resource use, impacts, and their monetizations into a distribution of predicted net benefits.

opportunity cost—the value of alternatives not chosen, calculated as part of an analysis of the costs of the alternative that was chosen.

quasi-experimental design—an experiment designed to produce evidence of causality when randomized controlled trials are not possible, using alternative statistical procedures to compensate for nonrandom factors.

randomized controlled trial—an experiment in which the participants are assigned by chance either to receive the intervention or treatment being studied or not to receive it, so the results can be compared across statistical identical groups. When this is done with a large enough number of participants, any differences among them that might influence their response to the treatment will be distributed evenly.

regression discontinuity design—a quasi-experimental analysis that can be used in program evaluation when randomized assignment is not feasible. It is based on the assumption that individuals who fall just above or below a cutoff point on a particular scale are likely to be similar, so that this group can be treated as varying randomly.

shadow price—the true value or cost of the results of a particular decision, as calculated when no market price is available; a dollar value attached to an opportunity cost.

standard error—used to refer to the standard deviation of various sample statistics such as the mean or median. The smaller the standard error, the more representative the sample will be of the overall population.

Appendix B

Workshop Agenda

Workshop Objective

The objective of the workshop is to highlight the issues on finding consensus on the standards for benefit-cost analysis of preventive interventions for children, youth, and families.

DAY 1: November 18, 2013

- 10:15 am–10:30 am

Welcome and Introduction
Jeanne Brooks-Gunn, Committee Chair,
Columbia University
- 10:30 am–11:30 am

Session 1: Overview of Benefit-Cost Analyses of Preventive Interventions for Children
Speakers:
Margaret Kuklinski, University of Washington,
Seattle
Charles Michalopoulos, MDRC

Session Moderator:
Janet Currie, Princeton University
- 11:30 am–12:45 pm

Session 2: Costing Interventions
Speakers:
Henry Levin, Columbia University
Max Crowley, Duke University

Session Moderator:
Jorge Delva, University of Michigan

1:45 pm–3:00 pm **Session 3: Valuing Outcomes of Intervention**
Speakers:
Steve Aos, Washington State Institute for Public Policy
Damon Jones, The Pennsylvania State University

Session Moderator:
Roseanne Flores, Hunter College, CUNY

3:15 pm–4:30 pm **Session 4: Standards of Rigor for Program Evaluations**
Speakers:
Rebecca Maynard, University of Pennsylvania
Thomas Cook, Northwestern University
Jens Ludwig, University of Chicago

Session Moderator:
J. David Hawkins, University of Washington, Seattle

4:30 pm–4:45 pm **Recap of Day 1**
Session Moderator:
Jeanne Brooks-Gunn, Committee Chair, Columbia University

DAY 2: November 19, 2013

9:00 am–10:15 am

Session 5: Other BCA Standards

Speakers:

*Richard Zerbe, University of Washington,
Seattle*

Anirban Basu, University of Washington, Seattle

Lynn Karoly, RAND Corporation

Session Moderator:

Melanie Lutenbacher, Vanderbilt University

10:30 am–11:45 am

**Session 6: Translating Results to Inform
Policy and Practice**

Speakers:

*Jacqueline Jones, Independent Consultant,
Formerly of U.S. Department of Education
and New Jersey Department of Education*

*Linda Smith, Administration for Children and
Families*

*Kathy Stack, U.S. Office of Management and
Budget*

Session Moderator:

*Gary VanLandingham, Pew Center on the
States*

12:00 pm–1:30 pm

**Session 7: Concluding Thoughts (Roundtable
Discussion)**

Session Moderator:

*Jeanne Brooks-Gunn, Committee Chair,
Columbia University*

Discussants:

Janet Currie, Princeton University

Jorge Delva, University of Michigan

Roseanne Flores, Hunter College, CUNY

*J. David Hawkins, University of Washington,
Seattle*

72

CONSIDERATIONS IN APPLYING BENEFIT-COST ANALYSIS

Melanie Lutenbacher, Vanderbilt University
Gary VanLandingham, Pew Center on the
States

1:30 pm

Adjourn Workshop Day 2