Advancing Social Policy-Making Through Benefit-Cost Analysis:

Challenges and Opportunities

June 24-25, 2008
The Liaison Capitol Hill
Washington, D.C.

Introductory Remarks—Michael Stegman

My thanks and appreciation to Dick Zerbe and his colleagues at the Evans School of Public Affairs at the University of Washington for doing the big and small things that are necessary to make a conference like this a reality—Brianna Hiser, Sharon Doyle, Georgine Yorgey, thank you all.

I also want to express my appreciation to a very important member of our policy team at MacArthur, Valery Chang, who is here today, and to our program assistant, Gerry Sims, who has worked with Val and me on our end to help make these next two days go as smoothly as possible.

In his closing remarks tomorrow, MacArthur president Jonathan Fanton will discuss his vision and hopes for the project that we call *The Power of Measuring Social Benefits*, so this morning I would just like to make a few framing remarks.

The MacArthur Foundation has an ambitious domestic grantmaking agenda that, among other areas, includes affordable housing and community economic development, mental health and juvenile justice, and a new $50 million initiative in digital media and learning that focuses on how children are using this new technology to learn inside and outside the classroom.

We are also making significant new investments in activities related to America's fiscal future out of concern that a failure of our political leaders to effectively address long-term challenges posed by the growing structural deficit could seriously reduce economic growth and our quality of life, doing disproportionate harm to the country's most vulnerable populations who are the focus of much of our grantmaking.

We have also become concerned that America's increasingly challenging fiscal condition has the potential to reinforce an undercurrent of popular and political opinion that often sees the interests of those in need at odds with the interests of the larger society. At the same time, we believe there is increasing evidence to the contrary, and if that evidence were to be developed further, and used effectively in the political process, it may be possible to challenge and perhaps shift that perception.

Over the course of this project that we call *The Power of Measuring Social Benefits*, we want to test the hypothesis that effective social policies that invest in individuals in need or at risk not only improve their life chances, but in many instances benefit the larger society and generate public returns long after assistance has ended.

Through a three-part grantmaking strategy, this multi-year, $35 million project is (1) supporting formal benefit-cost analyses in social policy domains of interest to the Foundation, and which seem to be ripest for policy advancement; (2) supporting intellectual efforts to strengthen methods and standards for the conduct of such studies, thus making them more useful for policymaking; and (3) working to increase demand for social benefit-cost analyses by policymakers.

It is the MacArthur Foundation's hope that through this strategy it will be possible to help build a compelling body of evidence that could reduce the role of ideology in social policymaking and accelerate the rate at which evidence-based practices are adopted across all levels of government. And in the process, it may help restore public confidence in government's ability to help solve critical social problems facing the country.

Although we have made more than $12 million in grants in support of this initiative over the past 18 months, this conference represents the first invitational gathering where our grantees will be able to discuss their work with thought leaders in the benefit-cost field, policymakers, government officials, and practitioners. From my standpoint, this is our first way of taking stock of where we are and gaining important insights into how we should proceed with this initiative in the future.

Again, thanks to Dick Zerbe and his colleagues for making this conference possible and to all of you who have found the agenda of sufficient interest to take your valuable time to participate in the various sessions over the next 36 hours.

Thank you.

The Use of Evidence-Based Research on Children's Outcomes to Promote Economic Competitiveness: The Case of the Partnership for America's Economic Success

Chair:
Robert H. Dugger is a managing partner for the Tudor Investment Corporation.

Panelists:
Greg Duncan is the Edwina S. Tarry Professor at the School of Education and Social Policy and Faculty Fellow at the Institute for Policy Research, both at Northwestern University.

Robert Bradham is the Senior Vice President of Business Development and Government Affairs for the Greater Richmond Chamber of Commerce.

Robert Dugger: Established in 2006, the Partnership for America's Economic Success (PAES) is a collaboration of 13 funders, business leaders, economists, and early childhood experts. The collaboration is managed and housed by Pew Charitable Trusts. PAES believes investments during early childhood can have significant economic returns. In 2007, policy successes included:

- The State of Illinois committed to Pre-K for all three and four year-olds, health care for all children, and increased fiscal resources for infant programs.
- The State of Ohio approved over $200 million for Pre-K, child care, mental health, and other interventions for young children.
- As part of its prison reform effort, the State of Texas raised home visiting funds by $7.9 million; $4.3 million of which was "re-invested" from the prison budget and will reach 2,000 new families.

The Partnership for America's Economic Success participates annually in the Telluride Summit, and has helped to develop and ratify the Telluride Principles, a common framework for successfully negotiating investments in early care and education. These principles assert that successful programs must have clear goals and measurement criteria, and rigorous and regular performance evaluations. Benefit-cost analysis can show that expenditures on human capital very early in life are investments that create long-term returns that greatly outweigh the costs.

Greg Duncan: Evidence increasingly indicates humans are very sensitive to stress early in life, and that it deeply affects behavioral, cognitive, and immune system development. The effects of early stresses remain throughout life, which suggests that early life may also be a time when people are particularly sensitive to economic conditions. Heckman's latest work modeling the production process establishes that early investments can have a synergistic effect with later investments.

A set of random assignment experiments in the 1990s evaluated the impact of various kinds of welfare-to-work policies on child achievement. The primary goal of these policies was getting mothers into the workforce, but some of the programs also worked to increase income and/or develop the mother's basic skill set. Child achievement was assessed using either the mother's report or the child's test scores. Programs that focused on investing in the mother's basic skills had small impacts on children's achievement; and policies that stressed getting mothers into the labor force as soon as possible also had a positive impact, though all but one of these impacts was statistically insignificant. In all of the above-described programs, total family income stayed constant and employment replaced welfare income. In contrast, programs that included earnings supplements in addition to focusing on getting women into the labor force *all* had statistically significant positive impacts. Each additional $1,000 of income resulted in a 0.06 standard deviation increase in child achievement. With the $3,000 per year impact of the Earned Income Tax Credits, the effect approaches one-fifth of a standard deviation, which if persistent, can have a significant impact on lifetime earnings for a child.

A specific example of these programs is New Hope in Milwaukee. New Hope impacted people living in two of the city's poorest neighborhoods in the late 1990s. Developed by activists, the program offered a set of benefits to anyone with an income under 150 percent of the poverty line who could prove they had worked 30 hours or more per week in the last month. These benefits were designed to make employment work better for families, and included an earnings supplement, childcare subsidy, health insurance subsidy (if needed), and temporary employment if they couldn't find employment on their own. The program was a bilateral partnership between staff and recipients, and was framed as a set of supports that should be available to low-wage working adults rather than a welfare program. The program had the enthusiastic support of business elites in Milwaukee and ran for three years.

MDRC carried out an evaluation of New Hope that included rigorous random assignment with a comparison between the New Hope program and a statewide welfare reform program in place at the time. Results show a modest but significant increase in income for New Hope participants of about $500 per year. Part of the reason this amount is small is that some people, such as those who were working two jobs, took advantage of the income support to cut back on the amount that they were working to spend more time with their families. Family income went up by about $1,300 per year and poverty was reduced by about 10 percentage points. For child impacts, there was a 0.25 standard deviation improvement according to teachers' reports, with a more than 0.33 standard deviation impact for boys. These results are comparable with Head Start impacts.

The benefit-cost analysis breaks down costs and benefits into program participant, taxpayer, and society portions. Calculations show a net cost to society of $1,220, which changes once child-related benefits to society are included, estimated at $1,300. Including child benefits, the benefit-cost analysis is essentially even. These numbers are somewhat speculative because the 12-year follow-up has not yet happened and that is the stage when costly youth behaviors tend to show themselves. The real wildcard is the impact of New Hope on the behavior of minority boys. Mark Cohen estimates the value of converting high-risk youth into non-risk youth at $2.4 million discounted to value at age 14. His estimate includes the costs of crime, lower earnings, and other effects. Using Cohen's values, just one in 16 New Hope boys would need to be converted from high- to no-risk for the program to completely cover taxpayer costs. Information is not yet available as to whether New Hope has been able to cover taxpayer costs, but early signs point in the right direction.

The Panel Study of Income Dynamics (PSID) was a longitudinal study following a representative sample of adults and children. The sample set consisted of children born between 1968–1975, all of whom could be followed until 2005–2006 (at which point they were between ages 30 and 37). This study enabled researchers to link early economic conditions to adult outcomes. A regression analysis looked for the effect of income during early childhood, controlling for birth-related and demographic variables, as well as for income throughout the rest of childhood. This analysis enabled researchers to look at the effect of income during early childhood. Outcomes measured included completed schooling, annual earnings, annual work hours and transfer income after age 25. The analysis also allowed for distinct effects of income increases for low-income and high-income families, allowing for non-linear effects of increasing income for low-income families.

In the regression, the simple correlations are as expected. However, to have a differential effect up to $25,000 and beyond (in the full regression showing income) also shows the remarkable impact of low income early in life. For example, a $10,000 increase in income every year from age 0–5 is associated with 500 more work hours later in life. If you take regression results, what you get is a $70,000 cost associated with $53,000 in personal benefits, mostly higher post-tax earnings, assuming that impacts

stop at age 37. Meanwhile taxpayers benefit by $11,000 per year. There was no evidence that the impact declines, and assuming that it continues until later years, the benefits well outweigh the costs.

This research suggests that income transfers be conditioned on the age of children, which could take the form of EITC supplements for young children or changes in income tax deductions for young children. In any case, the evidence supports a re-thinking of current income transfer programs, to focus on these particularly important early years.

Robert Bradham: Engaging the business community in early childhood issues is a powerful way to affect public policy. For example, 10 years ago, 60 percent of Richmond's third graders could not read at grade level. The business community was worried, knowing that this was their future workforce. Around the same time, Richmond hired a new superintendent to improve educational outcomes. In 10 years, the number of children who cannot read at grade level has been reduced to 24 percent.

The Greater Richmond Chamber of Commerce (GRCC) uses both public and private investment to do work in four areas: (1) affordable housing, (2) transportation, (3) education, and (4) where the workforce is coming from. The chamber uses benefit-cost analysis to help elected officials make that leap of faith and to "get beyond the warm fuzzies" to some hard numbers. Benefit-cost analysis studies are essential to persuading public officials, but they can only be selectively used, so that there are not a confusing number of different statistics flying around. Another challenge with using BCA is that officials often want more specific results, rather than broad ranges, or need to understand why the broad range exists. Another significant impediment to using benefit-cost data to inform policy decisions is a lack of local data, which would show the return in local communities

Despite these challenges, GRCC has effectively used brochures with faces of well-known respected individuals from the nonprofit and business communities, as well as police commissioners, school superintendents, and other community leaders. Inside the brochure, research statistics and benefit-cost analysis results get the message across. GRCC also had eight business leaders each do a 30-second video spot that was emailed to policymakers; the Secretary of Finance called to say this was the best email he had received all year. These efforts garnered $23 million for funding early childhood in a year with a $1 billion shortfall, when most other programs were being cut.

Q and A Session from Panel One

QUESTIONER: One thing that we're obligated to do as researchers is to avoid overstating the certainty with which we know things. I appreciate your need to convince lawmakers, but there is enormous uncertainty about all of this, and when we go to invest money, they don't guarantee a certain return. So part of this is about educating lawmakers that there is a certain amount of uncertainty in making investments.

In the research, I think that people often grossly overestimate the things that they know, and Professor Heckman's figure is one example of this. For one thing, if that graph had confidence intervals, we would see that the uncertainty dwarfs the differences that we see in rates of return. All the studies use different measures of costs, meaning that they aren't really comparable. The point of this is that it's fine to argue that early childhood can generate a good investment return, but is dangerous to suggest that early childhood should be done at the expense of early adolescence. It gives the suggestion that by the time kids get to adolescence they might be lost to us, ignoring the resilience that we see, for example, with crack babies.

DUGGER: First of all, you're exactly right. We look at uncertainty in many of these situations, but we also ask another question. In Virginia we have a budget that is a train wreck. Decisions are not made on the basis of uncertainty and statistics, but on lobbying power. The only uncertainty here is on whose lobbying power is greater. And this happens at the Congressional level as well. As long as it operates this way, Congress will be providing $300 million per year for the private jet industry. It's a tiny amount of the overall budget, but it would provide prenatal care to every mom in the country.

My instinct is that after you do the work that this conference is setting you up to do, at the end of the three year period, if you haven't pushed benefit-cost analysis, then the window of opportunity may be closed. If you don't answer these questions, the alternative is not careful benefit-cost analyses, but lobbying power. If you believe that decisions should be based on what we know of the facts, then you are obligated to push benefit-cost analysis as far as you can over the next 36 months.

QUESTIONER: As researchers, we also have clear standards about what is adequate quality for making decisions, and I would say that much of the economic analysis in this area of social policy does not match the research standards that exist in the area of health policy, for example.

BRADHAM: From the Chamber's perspective, we're not advocating investing in early childhood over adolescents. And on lobbying: Legislators don't seek me out and ask what the business community thinks about this. We have to seek them out, and when I do, I have to make the best argument that I can, using the best data that I've got, to compete with all the other interests that are out there. And that's why I'm using data that you provide. It helps move the business community and policymakers.

If we can get Dominion Power behind this, it makes a big difference, because they make a lot of political contributions. When a CFO of a fortune 500 company says he's in support of this kind of initiative, and if he says you'll get more back on your investment, it makes a big difference.

QUESTIONER: The reality is that there are multiple parties competing for children's dollars—whether education or health—so it's not only about determining which programs are most effective, but also about making sure that we are arguing for investing more overall.

Second, I'm also curious about whether any of you have been involved in the United Nations Convention on the Rights of Children. Given that there is lots of overlap with the things that you are talking about, have any of you looked at that process, to figure out why the United States did not ratify it?

DUGGER: To respond to your first question, the Telluride principles were drafted to provide context for the early education field to have a platform for agreement. To have an effective early education community, you first have to have that community agree on what is first priority. Second, families must be involved. And third, you have to agree that resources will be allocated on basis of best evidence. If these different elements within the early childhood field agree to use the best science to make funding decisions, then as evidence changes, so will allocation of resources. That agreement is the starting point for cooperation in this community—agreeing that there will be a lot of performance evaluation. As the benefit-cost community gets better at its job, we'll understand where resources should go. But this "declaration of peace" is an important first step to allow for cooperative collaboration.

As for second question of have we looked back—not really, in the sense that we think that there's pretty far to go between what the broader global community is doing, and where the United States fits.

However, those ideas were on the table at the Telluride meeting, and we moved forward. We will ratify a set of principles—not exactly those, but a set of principles.

QUESTIONER: Are you accounting for the cumulative effects of human capital?

DUNCAN: We were using Alan Krueger's estimates. We reviewed the literature on the causal relationship between academic achievement in high school and lifetime earnings, and that was the basis for our calculations.

QUESTIONER: Do you have people on staff who can do independent assessment of benefit-cost studies?

BRADHAM: We have someone on staff who analyzes research to evaluate their applicability to us, but not necessarily according to methodology or validity.

QUESTIONER: I was looking at Greg Duncan's calculations, and thinking about your proposal that we spend $4,000 per child, and thinking that there are 16–20 million kids in that category. At first I was thinking "how are we ever going to do that realistically?" But if you think about the annual increase in Medicare in Medicaid, we will spend more on those two programs over and above what we spend now per year than what we might spend on all these programs together. The idea that we could spend $1 billion a year on kids seems out of reach, but we spend that on increases in Medicare/Medicaid per year.

GREG: That question defines the problem we must grapple with: We don't have a process to disconnect the autopilot and then come up with a rational process for making these fiscal decisions.

Chair:
John Graham is Dean of the Pardee RAND Graduate School of Public Policy, and Dean-Designate of the School of Public and Environmental Affairs at Indiana University.

Panelists:
John Morrall is Branch Chief at the U.S. Office of Management and Budget.

Kimberly Thompson is an Associate Professor of Risk Analysis and Decision Science at the Harvard University School of Public Health.

Jonathan Wiener is Perkins Professor of Law and Environmental Policy at Duke University.

John Graham: There are significantly different types of benefit-cost analysis is currently in use, particularly in the ways that BCA is used by the U.S. government and within Europe.

1. **The Kaldor-Hicks Approach**. This is the traditional approach, distinguishing features include willingness to pay (WTP) as a measure of benefits, and opportunity costs as a construct for costs. A specific problem of this approach is how to deal with low income populations, whose willingness to accept (WTA) is likely to be large, but whose WTP is by definition going to be small.

2. **The Social Well-Being Approach.** In this approach, the unit of analysis is not money; rather, it is some interpersonal measure of well-being. Public health fields, for example, often use quality-adjusted life years, though there is also an entire literature about how to measure happiness and use it as a measure.

3. **The Social Welfare Function.** This approach uses an interpersonal measure of well-being, but also adds a measure of equity into the equation. It is typically theory-driven, but it the UK government has recently introduced distributional income weighting into their benefit-cost analysis.

4. **Social Risk Analysis.** This analysis recognizes that variability and uncertainty exist, and should be explicitly addressed and incorporated into the benefit-cost analysis. Accounting for variability and uncertainly is particularly important if you have vulnerable populations who are different from average people.

5. **The Multi-Objective Method/Regulatory Impact Analysis.** This approach is a portfolio of methods, which includes benefit-cost analysis among many other tools, and has no single normative foundation. This suite of methods creates a lot of discretion for analysts to choose which method and normative foundation(s) to use.

John Morral: To what extent does BCA improve decision-making? And wow can we increase its (presumably positive) impact with respect to the analysis of federal social programs?

Legal critics have placed Morrall and other similar BCA practitioners under the "Unicorns of Deregulation" umbrella; the practitioners include Bob Crandall and Bob Litan of Brookings, Bob Hahn of AEI, John Graham of RAND (and soon Indiana), and Kip Viscusi of Vanderbilt. The concern is that the criticisms are personal and not based on the merit and logic of their scholarly work, and public perception of BCA is being influenced by these legal advocates, who tend to be emotional in their appeal. Although the critics are mostly environmental law professors, they are influencing the acceptance of BCA in general.

There are legal scholars who have long defended BCA such as Sunstein and Judge Breyer, the more recent defenses have been written begrudgingly. For example, Revesz and Livermore's new book: *Retaking Rationality*: *How Cost-Benefit Analysis Can Better Protect the Environment and Our Health*, takes the approach that BCA is here to stay, so if you can't beat them join them. They recommend developing BCAs that will show more benefits for lower costs. They appear not to have carefully read most of the hundreds of BCAs that agencies have done in support of regulation over the last 15 years,

and they appear not to be familiar with Office of Management and Budget (OMB) Circular A-4, which advises agencies how to do BCA and deals straight on with the so-called fallacies of BCA.

Orrin Pickly in his June 2008 review of their book in *Science* magazine listed five of the "fallacies" of the way BCA is done in practice:
1. All unintended consequences are bad.
2. Older people are less valuable.
3. People cannot adapt.
4. Industry cannot adapt.
5. People value only what they use.

All of these issues are properly dealt with by Circular A-4 and agencies are well aware of them. A-4 is a state-of-the-art set of guidelines that has not gotten the U.S. recognition it deserves. It has been used by Canada and the European Commission in their recently issued BCA guidance. A-4 was peer reviewed and went through notice, comment, and interagency review. John Graham and his staff, with the help of the President's Council of Economic Advisers (CEA), specifically Randall Kroszner and John List, spent many hours reviewing the comments and revising the document. The document itself was an update of the 1996 OMB/CEA "Best Practices" document directed by Joe Stiglitz., former chair of the CEA. It contains, for example, the following responses to the aforementioned "fallacies:"
1. All unintended consequences are bad.
   →*A-4 has a section on "Ancillary Benefits"*
2. Older people are less valuable.
   →*A-4 explicitly recommends using a higher yearly life value for seniors (VSLY) because they face higher overall risks and enjoy higher savings*
3. People cannot adapt.
   →*A-4 recommends that for fairness and simplicity the population mean should be used so as not to undervalue the disabled when using QALYs for a disabled population.*
4. Industry cannot adapt.
   →*A-4 has a section on taking into account technical change and the learning curve.*
5. People value only what they use.
   →*A-4 states that not taking into account non-use value significantly understates benefits (p. 22). There is a whole section on Stated Preference Methods.*

Economists should not let the legal profession dictate how BCA should be performed, and need do a better job showing how BCA, or perhaps Cost-Effectiveness Analysis, has improved social decision-making by providing systematic and objective information. As mentioned, many of the critiques of BCA appear to have a simplistic view of how BCA is actually performed by the agencies and what the results of those BCAs indicate.

Since 1997, OMB has produced an annual *Report to Congress on the Costs and Benefits of Federal Regulations*. The Regulatory Right-to-Know Act requires OMB to report on costs and benefits in the aggregate, agency by agency, and by program and major rule. The record shows that from 1992 to 2007, according to the over 100 BCAs that have been done for major rules and reviewed by Office of Information and Regulatory Affairs under Executive Order 12866 issued in 1993 (Executive Order 12291 before 1993), the Federal Government has produced conservatively about $350 billion in annualized net benefits. Costs were about $120 billion and benefits four times larger at about $480 billion in 2007 dollars using midpoints of ranges. The handout shows the year-by-year costs and benefits. Note the pattern of large benefits during election years.

Kimberly Thompson: In 1988, the World Health Assembly (WHA) committed to eradicate polio by 2000, and when they failed to achieve this goal, several players in the discussion said that one contributing factor was that key decision makers had not done any economic (benefit-cost) analysis. They asked the Center for Disease Control (CDC) and its partners to help address this issue, and in the process, the group was able to help them think about the key concepts of uncertainty, variability, time, and system dynamics. To begin this work, they went back and looked at the historical costs and benefits in the U.S. of historical polio eradication, calculating total net benefits of $180 billion. Vaccines are one of the areas that historically generate huge net benefits.

What did the decision tree for global eradication look like? It was much more complex than the WHA originally thought: instead of the five options they originally saw, it actually included seven **categories** of options. This is a common outcome when you start to look at a global problem, because different countries may be using very different strategies, and are starting with different resources and economic and social situations.

To help them clarify this, they stratified the world according to World Bank country income levels first, then by background and existing conditions and risks, and then by costs. By combining all of this information, they were able to model different choices, and use those scenarios to make forecasts over the next 20 years, looking at the projected incidence of polio cases, and the net present value of costs. There were several highlights of the analysis:
- There will be outbreaks.
- There is not one dominant option.
- However, the option of controlling polio at low levels without eradicating it is a very costly option. From 2002–03, after eradication efforts reached 95%, counties scaled back on campaigns; 25 counties were subsequently re-infected and this cost $450 million more.

Based on this kind of analysis, they calculated that the world should be willing to pay $3.5 billion for eradicating polio in low-income counties alone. This type of policy analysis has also helped the World Health Organization and others to take a more long term perspective, and to resist being impacted by short term cost considerations

There has to be significant interaction between analysts and the people who own the problems. In dealing with complex global problems, we have had to fight against oversimplifying. In doing this, we are able to give policymakers tools to help them deal with and think about the complexity.

Jonathan Wiener: People frequently assume that Europe is more precautionary that the United States. This comparison is motivated by several high level issues on which Europe has been more precautionary than U.S, as well as their formal adoption of the precautionary principle: GMO foods, hormones in beef, climate change, toxic chemicals, guns, and anti-trust legislation. Media, scholars, high level officials, and policymakers all assume therefore that Europe prefers the precautionary principle to Benefit-Cost Analysis.

In evaluating these claims, the key is to distinguish anecdotes from social science, as anecdotes are not a representative sample. To take a more systematic approach, Wiener worked with others to launch the Reality of Precaution Project, which used two methods to make a better (if still not perfect) comparison. First, they took a much wider array of case studies, a method that is still vulnerable to sampling bias, but which gives more institutional and longitudinal detail. Second, they took an aggregate data approach,

which is broadly representative, but lacking in detail. When they looked at all the case studies, they found a number of counter-examples from the 1990s to the present in which the U.S. was more precautionary than Europe: BSE ("mad cow") in beef, air pollution, counter-terrorism, smoking, particulate diesel fumes, and youth violence.

Next, they used an aggregate data approach to test a larger sample, asking a large team of undergraduate and graduate research assistants to identify the risks mentioned in academic and government literature in both the U.S. and Europe from 1970 to 2004. They identified 11,086 different risks, combining those into a sample of 2,878 unique risks. Of those, they selected a random sub-sample of 100 risks; for each risk, they scored the relative precaution for the U.S. and Europe, for every year from 1970 to 2004. They then looked for trends in the data over time, both unweighted, and weighted by the severity of the risk.

They found a slight trend towards increased precaution in the E.U.; however the shift was only three to six percent and they found no significant overall difference between the U.S. and the E.U. Overall they concluded that there is no difference between the E. U. and the U.S. in relation to the precautionary principle. What they did see is that different societies worried more about some risks and less about others, leading them to ask why the U.S. and the E.U. worry different amounts about different risks, and how the E.U. and the U.S. are borrowing and learning from each other.

Since 2001, the E.U. has been pursuing an agenda of "better regulation," which has led researchers to look at how the E.U. has been borrowing and adapting U.S. ideas about regulatory policy and regulatory review.

In the U.S., versions of benefit-cost analysis have been used for a long time in many different areas of government, from public works projects such as dams, to welfare, tort law, and regulation. Congress has begun enacting laws that speak to benefit-cost analysis, but has not made up its mind: some of the laws require BCA, while others prohibit it. We do not yet have general approach to applying BCA to legislation.

There is a consensus among presidents since Carter that BCA is important to help manage the regulatory state, although there have been some important changes along the way. Reagan said that benefits must outweigh costs. Under Clinton, outweigh was replaced with "justify," to accommodate additional qualitative information. Explicit mention of weighing distributional effects and risk tradeoffs was also added. Under Bush, the Clinton executive order has remained in place, with some additions about reducing discount rates requiring public publicity analysis for very large projects and ex post review. Contrary to what some critics believe, BCA can say yes as well as no, and recently, prompt letters and routine administrative mechanisms are being investigated as methods to accelerate good ideas.

The E.U. has different set of institutions than the U.S. The European Commission, the executive body, has a monopoly on the initiation of new legislation. The European Parliament can respond to proposed legislation and can make amendments. There is also the Council of Ministers, another legislative body, made up by the heads of state of each member country. There are European Courts and agencies, including a little-known administrative ombudsman that has the power to bring action to remedy administrative misuses.

BCA has been introduced to the E.U. via a number of routes, including the proportionality principle (which the courts have interpreted to imply BCA for European Legislation), the precautionary principle,

and a requirement in the European Union Treaty that the potential benefits and costs of action or lack of action be evaluated. Many member states have laws requiring BCA, and the better regulation initiative also suggests BCA. Based on the proportionality principle, it is arguably a general principle of European law that BCA is required for new legislation. The precautionary principle is often viewed as opposed to BCA, but in February 2000, in a move that Wiener views as an attempt to reclaim the precautionary principle as decision analysis, the European Commission issued a Communication on the Precautionary Principle, which said that precautionary regulation must be proportional, and must be based on an examination of the potential benefits and costs, including, where appropriate, an economic benefit-cost analysis.

Under the Better Regulation Initiative (BRI), impact assessment guidelines have been adopted and revised, and the European Union has conduced over 100 extended, or full, BCAs, and focused on administrative cost reduction. Questions remain about how widely the BRI is being implemented, and a 2005 evaluation showed that some member states were very successful, while others were not at all. One popular aspect of BRI is administrative cost reduction, with a goal of cutting information collection costs by 25 percent by the year 2012. This campaign undoubtedly has benefits, but one weakness is that there is no accounting for the benefits of information collection requirement, which may lead them to make misinformed decisions to cut costs.

Major questions about the BRI include: Is there any judicial review and is there executive review? The courts have been fairly deferential, but in 2006, in a case about agricultural subsidies, the court ruled that the Council of the European Union was in violation of the proportionality principle in failing to conduct a thorough impact assessment. Originally, there was no executive oversight of impact assessments, but in 2006, the European Union launched a new Impact Assessment Board, which has the authority to review the quality of impact assessments, and the authority to prompt impact assessments, though it does not yet have the explicit authority to reject or prompt a proposed policy for based on the results of the impact assessment.

Lessons from the Unites States and the European Union:
- Seeking a full portfolio of important consequences should be prioritized above seeking a precise quantification of a narrow set of benefits and costs, and ignoring others.
- Do not focus narrowly on administrative costs.
- Pursue proportionate analysis—proportionate to the impact of the potential consequences, which means evaluating information costs and information impacts.
- Assess international as well as domestic impacts.
- Apply impact assessment of BCA more widely, rather than limiting it to risk regulation. This will help counter the perception that BCA is an anti-regulatory tool.
- Study differences in BCA methods between the European Union and the Unites States.
- Apply impact assessment (BCA) to legislation as well as rulemaking.
- Apply the proportionality principle to all lawmaking.
- Use BCA to prompt the good as well as to reject the bad.
- Establish a central expert oversight body.
- Learn over time, across agencies and countries.

Q and A Session from Panel Two

QUESTIONER: When I listen to these talks, I think one of the main issues is not whether you should be doing BCA, but what happens at the end? Is it a tool for decision making, or is BCA making the decision?

GRAHAM: I think the way to think about it is to ask: What is the nature of the constraint that BCA should put on a politician? How do you put some obligation on lawmakers to take analyses seriously? And this, I think, is a big problem.

WIENER: I think there is an acknowledgement that publicly elected or appointed officials, who are accountable to the people, have to make judgments. Part of the point of the word change (from "outweigh" to "justify") during the Clinton administration was to ensure that data informs judgment, but it's ultimately not going to supplant judgment.

THOMPSON: One of the things I've learned in the process of being an analyst working with decision makers is that we have to engage in dialogue with decision makers and be part of a team. The reality is that I'm doing part of the analysis, but I don't own the decision, and even if it's very clear who ultimately makes the decision, there are a lot of people who are providing information. This is a topic Richard Williams and I wrote a paper on; in the world of risk analysis, there are people who do the risk assessment, and hand it off to other people to do the benefit-cost analysis, without the context. We have to move more towards a team approach, so the separation between assessors and managers remains in place to avoid corruption, but the separation cannot be a complete divorce. If it is, you lose context and transparency.

Last thing is that I've learned that the reality of different stakeholders is really important, particularly in terms of variability. Stakeholders say, "Give me one number, but give me my number." We need to create the expectation that analysts are there to bring analytical rigor, tools, and objectivity; we are an investment in process, which is *more* than an answer.

MORRALL: I've been in the White House for close to 33 years and the big battle is about getting decision makers to even look at BCA, and not over whether we might blindly make a decision using benefit cost analysis. Originally BCA was a bipartisan effort, but now there's a bipartisan attack on BCA. We shouldn't just promote benefit-cost analysis because we think that might provide more funding for education or the environment; one example of an area that not been using benefit cost analysis recently is in the area of homeland security, as [BCA] has also been ignored by Republican administrations.

QUESTIONER: I worry that uncertainty and risk are being conflated, when they are distinct, and in fact, uncertainty is a much bigger problem analytically. Second, Circular A-4 is a much less detailed document than the EPA's document on preparing economic analyses, and I find this distinction very striking. There has been tremendous growth in the literature about social discounting, and OMB has stuck doggedly with a seven percent base rate, even though it also allows a three percent model. Is there any consideration to amending Circular A-4 in the coming years?

GRAHAM: The A-4 position shifted from asking people to use seven percent, to asking that you present both seven percent and three percent. This is a huge shift, for those of you who know the agency. For intergenerational questions, the guidance forbids agencies to use less than three percent.

RESPONDENT: Canada uses eight percent.

WIENER: The European Union uses a four percent discount.

THOMPSON: The way I'm defining risk is the probability of something happening, and we can be uncertain about the probability, but we can't then actually be uncertain about risk.

ZERBE: To go back to discount rates, we have to remember that it used to be 10 percent, with *no* possibility for adjustment. But I have a question as well, are there legislative changes that could be made that would enhance the proper use of benefit-cost analysis, whether it's through the OIRA or through some other agency, and related to this is the question of whether it's possible for OIRA to retain analytical independence over time, or is there some other arrangement that could or should be made?

GRAHAM: It's a huge topic, but one difference that Dr. Wiener pointed out is that the European Union applies BCA to legislation; whereas in the United States, we don't even get into benefit-cost analysis until after Congress has passed a law. I think the key question is: Can you bring something like the Congressional Budget Office into an analytic role around the original legislation? I don't know exactly how you would do it, but I think that would be a really powerful change. (Note from editor, the European system is used by the State of Washington. See Steve Aos' presentation in Panel Three.)

RESPONDENT: The U.S. does sometimes do BCA of legislative proposals. EPA did one of the Clear Skies, and it was a really dreadful document, and the Congressional Research Service exposed it for what it was. But I wonder if OIRA could have commented on it?

RESPONDENT: We did comment on it actively throughout the process. You probably would agree with some of the comments we made, and you wouldn't agree with others, but we were involved throughout the process.

Panel 3
Using Benefit-Cost Analysis in State and Local Government: Case Study of the Washington State Institute for Public Policy (WSIPP) and the City of Seattle

Chair:
Steve Aos is the Assistant Director of the Washington State Institute for Public Policy.

Panelists:
Tim Skeel is a Principle Economist for Seattle Public Utilities at the City of Seattle.

Steve Aos: The Washington State Institute for Public Policy (WSIPP) was created by the Washington State legislature in 1983 with the charge to carry out nonpartisan research on projects assigned by the legislature. Unlike similar institutions in other states, WSIPP does not seek federal funding because they work solely on projects of interest to the state legislature. The Institute performs various kinds of analyses, including BCA, and gives the results to members of the legislature.

In all BCA work, analysts and policymakers have to face the "real world" hurdle. Legislators are looking for conclusions, while benefit-cost analyses, on the other hand, deal with questions that generate testable hypotheses. WSIPP is trying to bridge the gap between researchers and legislators, and to make research findings work for legislators.

Over the last five years, WSIPP has been asked to produce a number of analyses that included BCA on topics as varied as prevention programs in early childhood, the K-12 education system, child abuse and out of home placement of children, substance abuse and mental health programs, developmental disabilities, crime and criminal justice costs, and prison construction. Recently, the legislature has become more comfortable with BCAs, and there has been much more of a legislative focus on program benefits and costs, fiscal impacts, and evidence.

When WSIPP is first assigned a BCA project, the first step is an honest assessment of the evidence about what works. There are two parts to this assessment: a systematic review of the literature, and either general findings or a specific list (which programs work or don't work), depending what's appropriate to the situation. Next, WSIPP performs two sets of benefit-cost analyses: a program-level analysis and a portfolio analysis. The program-level analysis looks at specific programs, reporting on net present value (NPV) or return on investment (ROI), or return on taxpayer equity. In contrast, a portfolio analysis looks at which range of activities the state should engage in to address a complex problem. This analysis is usually more complex, in part because it includes increased concerns about how complementary interventions interact with one another. Oftentimes, portfolio analyses are more useful to the legislature.

Between 1930 and 1980, adult incarceration rates in the United States and Washington state were steady at about two adults per 1,000 people. The 1980s saw a sharp rise in the incarceration rate. While crime rates have declined almost 25 percent since 1980, taxpayer costs have almost doubled. However, in December 2006 the incarceration rate was forecast to continue rising in Washington state, leading to a need for additional prisons. There is evidence that putting more people in prison has helped to lower the crime rate, but the best evidence from around Washington state and the U.S. is that increasing the average incarceration rate by 10 percent will lower the crime rate 3 percent at most. Legislators wanted to maintain the drop in crime, but were looking for lower-cost alternatives than incarceration.

WSIPP first looked at all research findings in Washington state and the rest of the U.S, and did a meta-analysis of 571 rigorous studies that addressed real-world applications. WSIPP then did a BCA,

estimating costs and benefits to the taxpayer and the crime victim in Washington state for each option. By applying the economic model consistently, WSIPP's study allowed legislators to compare different options easily.

The results, or the "consumer reports" lists use the same format to present what works, what doesn't, and what are the benefits and costs, and by keeping the format the same, whether the issue is crime, education, welfare, or health, legislators are more comfortable with these numbers, so that they are more likely to use the information. They may not exactly understand ROI, but they understand the icons and the color coding. What works is highlighted in green; what doesn't is highlighted in red.

So: what works? From looking at the program BCA, WSIPP could tell that adult drug courts made a difference. WSIPP looked at 57 drug outcome evaluations conducted in the U.S. that had rigorous research designs. On average, using some guesswork to account for selection bias, drug courts dropped recidivism rates by 8 percent. Without a drug court, an offender had a 58 percent chance of being reconvicted; with drug court, the chance drops to 54 percent—even that relatively small reduction generates $9,100 in taxpayer and crime victim benefits.

There are also other interventions that work, including cognitive-behavioral treatment, prison education programs and functional family therapy for juvenile offenders. For prevention, preschool programs and the nurse-family partnerships each have an impact. Notice as well that some of the programs didn't work. For example, increasing the number of contacts with parole officers didn't work in other states. Based on that evidence, the legislature was then able to move funding away from these types of programs in Washington state, shifting them into more impactful programs.

Finally, WSIPP went to the portfolio level, looking at costs of criminal justice options as a class. We developed three alternative spending models for the state—level, moderate, and aggressive—and made suggestions about the most economical ways to address crime under each funding scenario. For crime, this meant looking at prison supply and demand in Washington. Under the 2006 forecast, the state was looking at a two-prison shortfall by 2020 and a three-prison shortfall by 2030. The moderate funding approach leads to a need for fewer prisons, while the aggressive approach would lead to no new prisons required. Based on this analysis, the legislature funded a moderate approach that generated $1.7 billion in long-term benefits to taxpayers, with a benefit-cost ratio of $2.55.

The important thing to keep in mind is that legislators are real people who meet every year to make real decisions. People who are on the far left and the far right have no use for this research; however, most members in both parties care about these issues. Key lessons:

- Washington state's requirement of a balanced state budget forces choices, which increase the demand for BCA.
- Both parties would like to use taxpayer dollars wisely. This means that BCA has the potential to bring Republicans and Democrats together.
- The benefit-cost messenger matters, because trust is critical. There needs to be a local, long-term, and honest broker who presents this information to the legislature. We use the same standards for evidence every time, and use the same "consumer reports" format to present the information.
- To help people make decisions, you have to have consistent, comparative costs and benefits.
- Some specific techniques help us to give better information: meta-analysis (we gather as much information as we can, and don't play favorites); calculation of expected values and risk analysis of giving the wrong advice. These techniques allow us to say that we've looked at everything and that we're making suggestions that represent our best bet for what works.

- Big picture tangible outcomes matter more than ROI language. They say "I want to fight crime and lower taxpayer costs," not "what is the ROI on drug courts?" I need to know the ROI so I can give them a list, but they prefer the list.

As budgets get tight, in Washington as in other states, WSIPP will be able to continue to build on the relationships it has, and this will be good news for BCA.

Tim Skeel: On a policy level, as we provide more public goods through government, there is a danger that there are not enough incentives to economize. On the other hand, if most things are provided as private goods, there can be too many externalities. Each of these situations has a negative impact on social welfare. BCA, then, can help provide more public goods while keeping an incentive to create more social welfare with fewer dollars.

At the retail level, BCA aims to show how to reduce the amount of resources needed and improve the value of services we provide. However, at Seattle Public Utilities (SPU), we are not thinking only of the bottom line for our finances, because SPU isn't a private business. Instead, SPU looks at the "triple bottom line," meaning financial, social, and environmental impacts. This triple bottom line is the value SPU is trying to increase through BCA.

Six years ago, SPU decided that it should do a BCA for any expenditure with a life cycle cost of more than $250,000. To make that possible, SPU trained about 200 planners, public managers, and engineers in techniques for benefit-cost analysis, and instituted a process for doing BCA. Since then, SPU has done several hundred BCA analyses and tried to develop a template. Staff at many levels are asked to think about what they want to do, and then to think through the benefits and the costs. The idea is that BCA is not a hoop to go through, but a process to help you clarify what you want to do and the best way to get there. There are still some who see it only as a hoop, though. Some want to add all the benefits they can think of—even ones that are extremely speculative or attenuated—until a project is justified. Other times, projects are labeled such high priority that a BCA isn't necessary, only to have budget concerns later come into play. All of this is only to say that implementing BCA is not always going to be possible in an environment with conflicting interests, or when people are really convinced that they already have the answer.

In SPU's experience, the biggest benefits come from encouraging people to move from very little analysis to some. BCA is a way to keep people from making personal decisions, or doing things the way things have always been done. In these cases, even doing a little analysis can save a lot. There are fewer payoffs of moving from a thorough but relatively simple analysis to a very sophisticated analysis. Some barriers to effective BCA at SPU have included that people with vested interests are often put in change of analysis, that decision makers may have a personal analysis at odds with social BCA (triple bottom line), and that analysts may not have skills that are suited to BCA.

Q and A Session from Panel Three

SARAH WATSON, PEW: In the children's arena, outcomes are often softer and harder to put a monetary value on—more parents reading or softer discipline, for example, rather than waiting for improved third grade reading outcomes. How do these get dealt with by policymakers?

AOS: Those actually get well served by advocates. We get criticized because we can't put monetary values on all of those indicators; whereas, test scores are something we can put a value on. If there is

something that improves behavior and behavior improves test scores, and test scores improves lifetime earning, we can calculate that, but often the research doesn't yet exist for us to make that kind of translation. We get criticized when we can't measure it, and that's a fair criticism. We don't know any way around that, because we know that many programs that only measure intermediate outcomes don't work.

DAVID ZORN, FDA: How much time is there between a legislator's request and the submission of your report? How much staff time does it take?

AOS: We have a citizen legislature with a short (60 day) session one year and a long (105 day) session in the alternate years. Typically they want a request for one report and have it ready for the second session. Sometimes we can't do it in that length of time, and I have to tell them that. But part of it's being efficient, taking advantage of information that already exists. In Washington state, there is a lot of good info for BCA. You have to assemble it, but you can find it. We also try to assemble data in ways that makes it easy to update over time.

RICHARD BELZER: Often, problems that seem to be "no brainer" problems are actually more complex. And as analysts, one problem we have is that we subject results that we don't like to much more critical review than if we get something that we think is plausible. It would be helpful to be more systematic about this. In the WSIPP analysis of crime programs, I see five significant figures, and no treatment of uncertainty.

AOS: For the crime outcomes, those numbers are expected values of impact from the meta-analysis (for drug courts, 57 studies). We are 90 percent sure that this is different from 0, based on the studies I didn't show this today, but I actually do a sensitivity analysis, where among other things I vary the p value a little bit, and the effect sizes, and do a Monte Carlo simulation. After adding some uncertainty, I run the model 10,000 times, to check, given all my assumptions, what is the probability that there really is an effect?

QUESTIONER: You just used the term "sensitivity analysis," and I think that's something we should convey more often to policymakers. This is important not only to show how much uncertainty there is, but also to highlight that there are different effects across different populations, and that there may be high benefits for intervening with a specific group of people. Where we don't see an effect, it's not usually variability, but because the intervention is actually different, or because the populations are different. What do legislators ask you about?

AOS: Those are very sophisticated questions, and I don't usually get them. They really turn it over to us to help look at the language in the legislation. Part of the trick is to make sure that the implementation supports the research, and the legislation embeds directives and tools to target funding to the right population.

QUESTIONER: You said that you wanted to spend taxpayer money in the best way, and yet it seemed from your stories that you were actually trying to just ensure that there were equal benefits to the money being spent. If you were really trying to spend taxpayer money in the best way, I would think that you would want to put together a portfolio of the best programs, even if some of them would up having negative net benefits. In trying to be conservative, you are biasing the analysis. Are you aware of this?

AOS: We want to do the best thing that we can with whatever group we've got. Sometimes we have a lot of evidence, and sometimes we don't have much. With crime, we have a number of things that appear to be above 1 (a benefit-cost ratio > 1) that serve most of the populations that we need to serve in that particular market. So that requirement makes sense in this case, because we have enough comparative information.

Panel 4
Integrated Administrative Data Systems: Generating Benefits and Costs in Real Time over Time

Chair:
Dennis Culhane is a Professor of Social Welfare Policy at the School of Social Policy and Practice, and a Professor of Psychology in Psychiatry at the School of Medicine, both at the University of Pennsylvania.

Panelists:
Richard Burgis is Manager for Database Area Supporting the Department of Health Statistics and CSES, with the Department of Information Technology for the State of Michigan.

Pete Bailey is Chief of the Office of Research and Statistics for the State of South Carolina.

Manuel Moreno is the Research Director for the Chief Executive Office, Research and Evaluation, for Los Angeles County.

John Fantuzzo is the Greenfield Professor of Human Relations at the University of Pennsylvania.

Martha Moorehouse is Division Director for Children and Youth Policy at the U.S. Department of Health and Human Services.

Denis Culhane introduced the concept of integrated administrative data systems—tools that allow for the analysis of a broad range of projects (rather than a single project), over time. Because of the range of data collected, they improve understanding of policy impacts, and allow for analysis in "real time."

Culhane became involved in this emerging field through an interest in homelessness. Advocates have long believed that it was less expensive to house homeless people than to have them on the street, but there was no data to support this. When New York City built almost 5,000 units of housing for homeless people with mental illness, it provided a unique opportunity to gather data to test this proposition. Over eight years the 5,000 people who moved into the housing units and could be compared to approximately 8,000 people who didn't get into the housing and served as a control. His group started their analysis in 1996, and because the project was well underway, they planned to use existing administrative data. It took five years for him get the various agencies to share their data, which led him to support an administrative infrastructure that allows for continued data sharing on an ongoing basis.

His group found was that the average mentally ill homeless person in New York City cost the city $40,000 in services in 1999 dollars (and after all that expense, still living in a cardboard box, which is not a very good return on the taxpayer's investment). Individuals who were housed used $16,200 less each year in services, roughly equal to the cost of the housing intervention. From a purely economic point of view, about 95% of the costs of the intervention were recouped in decreased use of services.

Michigan's Statewide Data Warehouse

Richard Burgis and his team started working on an administrative data system for the State of Michigan in 1993. Initially they built a data warehouse within the Department of Social Services with the goal of reviewing the utilization of Medicaid and surveying for fraud. When the system was up and running, the governor decided to re-organize the Department of Social Services into two new departments: the Department of Human Services, which housed everything but the medical programs; and the Department of Community Health, which housed all the health programs, including mental health services and public health—both of which had previously been housed in different agencies. This reorganization provided a great opportunity for the data warehouse. They were able to help Community Health combine several different data systems from programs that had previously been housed in different agencies, and it allowed them to show their worth.

In 2000, once the resource was fully operational, they began to investigate how people in the state could use the data. They talked to policy people in the state about the problems that they were facing, and trained them to use the system. They now have more than 9,000 users and are one of the biggest,

most widely used data warehouses in the nation. People who use the database work in all parts of the government, and the data and resulting analyses have saved the state upwards of $500 million. The system currently includes data from eight different agencies; the big three agencies are housing, community health, and the state courts.

There is a significant amount of operational information in the data warehouse, which helps agencies cut costs, and is also used for audit work. For example, in the case of the courts, there are 83 different court systems (that don't tend to communicate with one another), and with an integrated data system they have access to events in other jurisdictions, such as whether someone has an unpaid ticket. There are also data-sharing agreements, which allow agencies to share information. For example, the Departments of Human Services and Community Health can to use corrections data to get a better handle on what the actual costs to the state are. Meanwhile, corrections can figure out what happens to people who leave corrections facilities and are out in the community. In another example, the state immunization program uses the program to find children and youth who are missing immunizations, from birth records, or from records of youth and family enrollments in other programs. Data-sharing also makes it easier to find the medical records of children in foster care, as children move from family to family, thus improving medical care of children in foster care.

In addition to the main data warehouse, there is a Statewide Homeless Assistance Data online Warehouse (SHADoW), which matches information for homeless families across human services, community health, and homeless agencies. After information is matched for individuals, they remove all identifying information, which provides researchers with a database where they can analyze data longitudinally, or conduct cost-based studies. For example, they can look at the events that lead up to homelessness, the costs incurred once a family becomes homeless, or the costs, and effects, of any interventions in which a family is involved.

The system is designed to be responsive to the people who are using it, and help them to do a better job. Not only do they see the results in the quality of their work, but it has also produced significant savings for the state.

Administrative Data Integration and Policy Analysis in Los Angeles County: The Adult Linkages Project

Manuel Moreno spoke about the Adult Linkages Project (ALP), which has worked over the last 10 years to link social service data for indigent adults and has greatly facilitated policy analysis within the county government. Two years ago, a Los Angeles county official asked Moreno's group to help the county better understand the General Relief program, a county-administered program that serves indigent adults who are ineligible or waiting for federal support. The challenge with the program is that people who are served by the program receive services from several different agencies, but the data was not originally shared between agencies.

The ALP produces information on service utilization patterns, gaps, and data before, during, and after individuals participate in Los Angeles County's General Relief (GR) program for indigent adults. The overall objective is to provide policymakers with empirical data to improve social policy in LA.

One of the challenges they faced is that some of the data is protected by privacy laws, so they wanted to de-link personal data from analysis, which allows Los Angeles to comply with privacy laws and to develop data that is sufficient for policy analysis. Nine county departments have singed memoranda of understanding (MOUs) with the warehouse; however, addressing privacy laws has been a complex

process, including extensive work with the County Council. It has taken over a year to explain how the data will be stored and that individuals' information will not be identifiable.

The ALP Data Warehouse links the administrative files of nine service agencies: Health Services, Mental Health, Public Health, Children and Family Services, Community and Senior Services, the Probation Department, the Sheriff, and Public Social Services. Historical data covers the types of services delivered, the length and timing of services, the costs of service delivery, the spatial distribution of service delivery, diagnostic codes, and patterns of multiple service utilizations. Data pertaining to a given individual is stored under a random project ID, which allows for analysis without identifying any client personally. The nine agencies covered by the MOU have access to the data, and outside researchers must obtain approval from the Chief Executive Office and the County agencies that signed the MOU.

Because this is a demonstration project, they are hoping that the Adult Linkages Project will provide a backbone for the eventual design of systems that will provide services in a more coordinated and effective manner. For example, they recently completed a benefit-cost analysis of a housing rental subsidy (part of a larger Los Angeles County Homeless Prevention Initiative) and because of the data warehouse, they were able to complete the analysis within three or four weeks and generate a preliminary report right away. The warehouse enables them to be much more responsive, and helps the programs' supervisors to improve cost effectiveness. Because of the success of the demonstration project, they hope to expand the data warehouse, with the next phase focusing on service delivery info for children and families.

Kids' Integrated Data System (KIDS): Integrated Data Systems Serving Children and Youth in Philadelphia

Over the last ten years **John Fantuzzo** has been working with the KIDS group to develop an integrated data system in Philadelphia on children and youth. KIDS is a partnership between the city, the school district, the University of Pennsylvania, and the William Penn Foundation, a local funder. The involvement of William Penn involved allowed them to develop a system that was independent of a particular administration or any particular group member, which has ultimately made the system function better. The KIDS system includes data from 1990 to the present and integrates municipal records for children and youth ages 0–21. It has the capacity for integration at the individual and aggregate level, including Geographic Information Systems (GIS) capacity for all data in the system.

Creating an MOU for each of the seven different agencies contributing data to the project was a time consuming process; however, once it was done, each individual researcher could focus on research rather than having to go to the city lawyers on their own to gain access. The entities contributing data include: the school district, the Department of Human Services, Vital Statistics, the Department of Public Health, the Department of Behavioral Health, Family Court, and the Office of Supportive Housing.

The data system is overseen by the KIDS Policy Group, which is made up of a representative from each agency that has contributed data; and the group works together to set up a research and review process. It was critical to the success of their efforts that each representative had the capacity to act for their agency. The research and review process itself was also important, because in order to comply with the Health Insurance Portability and Accountability Act (HIPAA) and the Family Educational Rights and Privacy Act (FERPA), you have to indicate that there is a significant administrative use for the research that justifies data-sharing. In fact, the policy group addresses all the barriers and challenges

that come up in using the data: legal, ethical, scientific, and economic. The legal requirements (HIPAA and FERPA), may seem like they unnecessarily keep people from using data; however, there are provisions in the law to allow for de-identified, or aggregated data, and they hope that their MOU will serve as a model for others who are trying to do similar work.

Ethical issues were, and remain, the most difficult, which include fears agencies have about what external researchers will do with their data. The group had to ensure that agencies never lost power over data and administration of data, and worked to set up a research system that created more benefits for them than costs.

They also had to deal with the issue of data quality. They needed to ensure that any data included in the system would be scientifically rigorous, and identify standards so they could identify what data was "bad," and to help agencies think about how they could improve it if they wanted that data in their system. (They didn't want efficient access to bad data.) And finally, they wanted a system that would support itself, and that would sustain a high level of use, independently.

The KIDS Policy Group has a process for screening projects and interacting with researchers. A researcher has to have an agency that is willing to champion their research, which ensures that projects only go forward when agencies have the capacity, or are ready to deal with, the outcomes of the research. The Policy Group reviews the proposals for scientific quality (this is separate from review by the researchers' own Institutional Review Board, which is also required). They also monitor the projects along the way, and are very attentive to the issue of timeliness—they want results soon enough to act on them, rather than five years down the road. There is also a requirement to present findings to the agency, and advance review of papers. Finally, there is a process for returning data sets to the Policy Group so they can be destroyed.

Examples of projects that have gone through KIDS include: a study looking at the impact of early education and kindergarten on school readiness; a longitudinal study showing kids' progress from when they were small children; and a study looking at children who co-occur in the child welfare and homelessness systems, and the timing and involvement of both systems. One of the biggest benefits of the system is that it creates a dialogue or a place for dynamic tension and conversations between researchers and government, which is very fruitful in integrating these two different groups.

Integrated Administrative Data Systems: South Carolina Office of Research and Statistics

Pete Bailey spoke about his experiences with the South Carolina Office of Research and Statistics, which received a grant in 1992 from the Robert Wood Johnson Foundation. During the planning year they carried out a major assessment of 450 public, private, and nonprofit agencies and organizations about key policy issues. They asked the agencies what data they would want if they got a wish from the "data fairy," surveyed what data was available, and looked at what data they would need to answer the questions that people wanted to ask.

They found that people needed to be able to link to other data systems to answer their questions, which meant that they would need to create person–to-person links between systems. At this time in South Carolina, there were strong business coalitions, and they were very supportive of vertical integration, which allowed them to get the project through the legislature without anyone raising objections (though it also helped that they weren't asking for money).

They started the project by adding identifiers to the records in the Health and Human Services (HHS) agencies and working with extracts of the data from legacy systems. In the early days of the project they were the information systems contractor for Healthy Start, they were able to turn Medicaid claims into data and incorporate it into the integrated system, and integrate all of this information into school report cards. This integrated information allowed them to analyze the data to find out new information. For example, they knew that there were disparities in the education system, and they thought it was due to poverty, but once they had made the linkages with the Medicaid data, they were able to see that the disparities had to do with race.

Today's South Carolina Integrated HHS system emerged from this early work. Building on existing systems, they have separated identifiers from statistical data and all data in the system is "owned" by the originating agencies. Other agencies must have permissions to use and link to another agency's data—this permissions system has paid off enormously. Records are linked using a unique tracking number that is randomly assigned. They can link and track people across systems, and they have addresses to do geocoding. The more data that they have in the system, the more that people want to be able to access it. If you pass your data to them, you stay in control of data, and you are linkable to any system. You lose nothing and gain everything.

Rules for use:
1. State agencies and nonprofits can control the use of their data, and they have to approve any use, or "linking" of their data with other data.
2. Mandated data (designated by the state legislature) is controlled by the Data Oversight Council.
3. Standardized applications facilitate use of, and release of, data. The Office of Research Support (ORS) coordinates applications and assists in the process
4. *Every* requestor has equal access, even ORS, unless specified by law. This means that there is no "owner" of the system who can control what analysis can be done.
5. Treat data with utmost confidentiality and a secure IT environment (HIPAA-compliant).
6. Over the long run, this environment encourages trust, and facilitates policy-relevant analysis and research

Even today this system is prompting new kinds of analysis as analysts can look at particular questions that had not been accessible to them previously. For example, they are doing a longitudinal analysis from data going back to 1995 to study diabetes; and they have been able to add geo-coding to data, so that local officials can figure out what problems are important in their area. It has also allowed them to dig more deeply into the determinants of success in schools by linking data detailing students' performance in school with health data (Medicaid data, detailing kids' health issues, or even prenatal issues), social problems, behavioral health, juvenile justice, geographic variables, and school level data. Researchers who are interested in children with ADD can sort the database to look at only these children, and find out how these children became a part of the system in the first place.

This integrated data has allowed the state to push people to use data to look objectively at problems, including performing more benefit-cost analyses (and paving the road to more BCA in the future).

Integrated Administrative Data from the Perspective of the U.S. Department of Health and Human Services

It's not an easy process to put together integrated data systems, notes **Martha Moorehouse**, and to do it in even five years is incredibly fast. In the end, it is the *process* that matters—you can only get access

to the data through building the relationships, and having the will behind the purpose. This process highlights the question of what we are trying to do with government spending and leads us to ask how we are doing, across the range of services that we provide, rather than silo by silo.

The U.S. Department of Health and Human Services (DHHS) has sponsored some of this work in the past, as we wanted to highlight states that were building this capacity, and also encourage states to look at the well-being of kids in their states, to ask what measures of well-being they were trying to move, and how they would know whether they had achieved their goals, using a range of data strategies, including their administrative data sets.

At that time, Vermont was the leader (as far as we knew) in this kind of work, and provided the form for others interested in developing this capacity. If you are interested in developing this kind of system, peer networking could not be more important. Every state needs to find its own way through the relationships and the legal issues, but at the same time, the more that you are armed with the experiences of others, the more you can effectively talk to lawyers who say that it is impossible, and the better your chances of success.

For DHHS, the data speaks to the need for understanding where our dollars are being spent. It allows us to see that different services are trying to serve overlapping populations of people, so we can start to focus on the concept of service pathways—how do the people come into the system, and how do they exit, and to ask if this a cheaper, or more expensive outcome; a socially better, or a socially worse outcome. Once you ask these questions, you can begin to ask which array of services are best to serve these people and how much it will cost.

Although, when you factor in evidence-based programs, it can be challenging to put this tool together with integrated data systems, and use them both to make decisions about policies. People immediately want to do benefit-cost analysis, but we often don't know that much about what costs are involved. And meanwhile, it takes a while to get the impact data that allows you to monetize benefits—all of this means that these results aren't immediately available.

In the meantime, what are the change strategies that DHHS can implement to help to create benefits for example, kids? When we work with programs to figure out how to improve publicly funded early childhood education, there is a huge difference between the "hothouse" programs that are the subject of evidence-based analyses, and the publicly funded programs. The majority of program developers in this field come from a background of child development, and they are not thinking about costs, so DHHS is spending time to help turn them into economists, to look at what the basic programs cost, and to look at what the enhanced projects cost. It would be great if they had monetized benefits earlier, but they can't even get to that point, if they don't know what it costs.

When looking at the transitions of kids coming out of foster care, and the transitions to adulthood, we have seen several trends: many of these kids are doing less well than their counterparts who are from low-income families nationally. But in addition, there are some kids who connect early to work and fall out, some who work early and stay working, and there are some who connect with work later on. This suggests a variety of different possible intervention strategies. Each time you we do this data work, we see the gaps—we don't know about kids who have left for the military, or kids who end up in jail. Each time we do these analyses we are pulled in the direction of getting more data and looking further. Cracking open the data and seeing how kids are actually moving through services can also suggest new policy directions to pursue.

Q and A Session for Panel Four

CULHANE: This discussion makes me realize that a lot of these systems may want to look at income data, and they don't all have access. For example, in SC, you don't have links to revenue data, do you?

BAILEY: When we looked at the data available from the Department of Revenue, we didn't like their data. But we are moving towards linking some of the educational data with commerce data, and if we can pull that off, then obviously we'll be moving forward.

CULHANE: Many of these outcomes are obviously related to income, and in Pennsylvania, we're now almost at a point where we have 18 years of data, so we can follow kids from birth through high school graduation. We're looking forward to being able to look at this data. Meanwhile, while we can currently link to a lot of negative outcomes such as juvenile justice records, we're also looking forward to the ability to link to positive outcomes, such as higher education data and employment information.

QUESTIONER: We think of this integrated data as our data, but in fact, another way to look at this is that it is a huge marketplace. There are millions of web users out there who want to use this data to hold government accountable. If we can make this data available to them, we can magnify the power of the data to actually change our government.

QUESTIONER: I may not be listening carefully enough but I hear a contradiction. On one hand, agencies have ownership and retain the right to sign-off, but there is also the issue of openness. Is anyone making sure that agencies are not improperly restricting access?

FANTUZZO: For us, we actually broker the contact so that the individual researcher doesn't have to make relationships with each agency. One dynamic that encourages more access to the data is championing access to each other's data, which encourages good projects to get done. This creates a dialogue that helps to surface, and publicize, what the highest priority projects for the administrative data are, and encourages researchers to focus on the issues that are most important.

BAILEY: I can't remember an agency turning down a request, but your point is fairly given. With agencies, access to data is open except to the point where an agency is using another agency's data to go to the legislature, the press, or for a grant—at this point we think that they should have to get permission, to encourage collaboration. So far, we haven't seen agencies who are trying to use the data in this way, but we think that it's important to encourage the process to be collaborative.

STEGMAN: Can you say more about the public use of data and the need to meet the requirements of HIPAA and FERPA?

CULHANE: If the data is de-identified, then you technically can have public access. But in Philadelphia, we don't have that kind of access. Instead, we require that there be an administrative use to the program. In the case of FERPA, it has to have instructional value, and for HIPAA, you have to have a value to the health agency. We observe this, even though technically it's not required.

MORENO: Most of the studies that are currently being planned are going to be used by the agencies. Over the years, we have developed collaborations with local universities and made our work well-known by the academic community. We haven't come across this question in our work. I think the issue of

openness versus restriction would be addressed by the agency that's deciding whether or not to share their data.

BURGIS: Making the data public is going to be a difficult task. If it's going to happen, it'll have to be a legislative mandate. De-identifying data is harder than it looks. There is a lot of publicly available data, and private data, that people have in their possession, and I think that it's going to be hard to make this data publicly available. It's certainly something that we should do, and if we can figure out how to technically make it happen, we could do it. But then on the other hand, there's the question of big brother.

MOOREHOUSE: General accountability questions that I see being addressed come out of different data sets and they have more to do with other areas. For example, the data can show that the money is not being spent on the services on which it is supposed to be spent. When kids are in systems and aren't making the progress that you want—this is a different question. Many of the agencies have citizen boards of some kind, which allow for some citizen input, and they would have the opportunity to weigh in on these kinds of questions, if they have access to the data.

RESPONSE: I think of this as a generational question. Here you have all this data that you are linking, and suddenly you can ask new questions that no one ever thought to ask before. And if you could put it on the web, and make it available to a whole new group of people, they're going to think of questions and connections that no one ever thought of before.

QUESTIONER: The Toxic Release Inventory is interesting because Congress requires it to be released to the public. But it was then misused—in a sense, it's designed for abuse. What I'm concerned about is the proprietary nature of the data. My sense is that if the pubic had any clue about what you were doing, they would shut you down in a heartbeat and the only way around that is doing something that is much more public and transparent.

CULHANE: It is a question that's related to the ethical guidelines that govern the use of the data.

BAILEY: It also is a question of encouraging good researchers. We will never get out of the hole we're in, in SC, without fantastic researchers, but even researchers have trouble understanding the datasets because they are complicated. What we would like in the House and Senate levels is to have the data digested, so people can make simple queries.

QUESTIONER: This is the vanguard and it's very hard to get state and local governments to pool data like this. In my experience, the issues are related to statutory and regulatory restrictions that are supposed to protect privacy. Are there tools out there to help us address those concerns when they are raised?

FANTUZZO: It's important to have some productive push back from agencies because this helps us to understand how they are trying to protect themselves and to benefit from the data. It is important to have someone else there who really understands what the privacy issues are, and it is important to have someone forcing due diligence of the privacy issues and creating guidelines. Understanding institutional narcissism is a great place to create dialogue.

KIM THOMPSON: Two comments: One is that it is remarkable that we have not had any mention of the National Children's Study; and two is that no one has mentioned Kids Count, the Casey Foundation initiative, which has been conducted for a very long time, and which gives us some national numbers.

CULHANE: As far as the National Children's Study, it has cost $1 billion, and I have been a critic of this effort, because if you were to take this money to the state level, you could have a study in every state, on every kid across the country, on an ongoing basis. Meanwhile, they're taking just a subsample, and doing primary data collection, which is very expensive. Yes, we should be doing data collection, but we should really be investing in creating the infrastructure for data collection that already exists and is often already standardized because of federal law. I see the National Children's data as an out-of-date approach.

To address your second point, Kids Count is aggregate data, but it would be interesting to see Kids Count variables that involved the linkage of two data sets together.

MOOREHOUSE: When we work with states that are using these administrative data systems, we often see that the same groups are looking at Kids Count data and blending it with administrative data to do the analysis.

CULHANE: I would still come back to the question of whether there are instances where they are integrating the data into complex, multi-agency questions.

BAILEY: Kids Count is a snapshot, whereas here we're talking about the same individuals over time, so it's sort of a different question. But what I'd like to emphasize is the unique contribution that the administrative data systems can make: By making the linkages between different kinds of administrative data, we're trying to break down the barriers that keep social workers from seeing only social work-related questions and to foster communication across these various agencies. If we can do that, then we are really starting to harness the power of these systems.

Panel 5
Key Factors Enabling Rigorous Research to Influence Policy: Lessons from Welfare, Education, and Other Areas

Chair:
Jon Baron is Executive Director of the Coalition for Evidence-Based Policy.

Panelists:
Robert Shea is Associate Director for OMB Administration and Government Performance at the Office of Management and Budget.

Ron Haskins is a Senior Fellow and Co-Director of the Center for Children and Families at the Brookings Institution.

John Wallace is Vice-President at MDRC.

Robert Slavin is Director of the Center for Research and Reform in Education at The Johns Hopkins University, Director of the Institute for Effective Education at the University of York, and Co-Founder and Chairman of the Success for All Foundation.

Robert Shea presented on the Office of Management and Budget's (OMB) collaboration to create a program assessment rating tool that determines the quality of management and performance of a program. This collaboration, initiated in 2001, consists of a set of basic questions one could use to ascertain the quality of management and performance.

OMB asks programs the set of questions and then tallies points—50 percent of the rating depends on whether a program has clearly stated goals, and the remaining 50 percent is based on whether or not they have met those goals. Over the last six years, OMB has assessed more than 1,000 programs representing the vast majority of mandatory and discretionary spending in the federal budget.

When OMB started assessing federal programs, over half of all programs were rated as ineffective or failing to demonstrate results. Now the number of ineffective or failing programs is down to 20 percent, representing a vast improvement in the quality of definitions of success. The process is rigorous, and includes a collaborative assessment and an auditing and appeals process. All evidence is available to the public and posted on the internet.

Despite this progress, the standards used to assess whether or not an agency has performed a rigorous, independent evaluation of its outcomes are fairly soft. OMB asks them to use the most rigorous evaluation practical, but very few programs have evaluations that are rigorous enough to show that the program itself, independent of other factors, is what is achieving the program's goals. For example, after the Secretary of Education inventoried several hundred math and science programs, OMB inventoried them, and found that only one-third (115) of the programs had carried out evaluations they felt were rigorous enough. Of those, just 15 had conducted impact evaluations that were rigorous enough to isolate the effects of the program. Three of those showed that the program was having a positive impact.

Shea is optimistic about these numbers, as he views those three positive programs as pieces of evidence that decision makers did not have access to previously, which can now be used to help make decisions. Similar examples of evidence-supported improvements have been seen in other areas, including healthcare, and the availability of this evidence is exciting progress. The next challenge is implementation, to make sure that programs are implemented correctly and that impacts are measured effectively.

Ron Haskin talked about how OMB's tool has the potential to help determine which programs are effective. He is concerned, however, that people perceive that this is part of a larger partisan plot to cut social programs. Even if the results are sometimes hard to take, because researchers know a lot about problems, technology, and methods to evaluate effectiveness, research should play a greater role in

policy formulation. Haskin noted it will be interesting to see what will happen with the next Congress and president's administration.

Today most of what determines policy are the media, the current administration, lobbying, and politics; in short, not research. While some of these elements are unavoidable and necessary, Brookings' goal is to expand the piece of the pie that draws on research, so that policymakers understand the problem better and can devise better solutions.

A great example of policy being shaped by research and results is welfare reform over recent decades. A study in the 1980s showed that there were two broad groups of welfare recipients. One group used welfare for short amounts of time; whereas, the other group stayed on for very long periods of times, and went back on welfare even if they went off. This type of evidence significantly changed the welfare discussion, because it showed that welfare dependency was real. But here is the other crucial part— when welfare reforms happened, assessment was built in. Results from these assessments in the mid-1990s showed that roughly speaking, if you had a program that emphasized helping people to find a job, these programs had positive impact. Since then, there has been a great increase among the numbers of people working, especially unmarried mothers. There was also an increase in both welfare and non-welfare earnings. Haskin asserts that even though the reforms that were imposed *were* draconian, the limits and particularly the sanctions produced results, and these general contours were shaped by research.

There are several lessons here: Yes, research helped define the problem and the solutions, but these results were also consistent with the political philosophy and the general legislation that had already been written at the time the research came out. If this had not been true, it's likely that the research wouldn't have had any impact at all. Even recognizing that the political environment is extremely important, Haskin believes it is researchers' job to bombard Capitol Hill and the administration with solid research. Testimony, work with the Government Accountability Office and Congressional Research Service, and particularly personal relationships are all parts of an effective strategy. Research organizations, especially large ones, need to develop these kinds of relationships and have a strong communications strategy.

John Wallace spoke about research's indirect influences on policy. Evidence, research, and BCA can have an influence on policy through impacting policymakers' beliefs. Negative impacts, or no impact can also influence policymaking, and small, positive findings can have a seemingly disproportionate effect. Information about long-term impact also establishes a baseline against which future programs can be judged, creating a knowledge base that builds off of, and goes beyond itself. And throughout all of this, methodology matters. Rigor, reliability, and transparency are all important. Random assignment designs enjoy a distinct credibility advantage and are increasingly required. It's easy to grasp and results are clear and unambiguous findings.

An example of these influences can be seen in changes in welfare program policies. Surveys indicated that welfare recipients themselves felt it was fair to require something in return for their receiving welfare, and these results changed attitudes and beliefs about welfare was. In addition, evidence that upwards of 50 percent of welfare recipients were already working helped change attitudes as well. Results of benefit-cost studies also had a marked effect. Research showed that from a fiscal perspective, it made sense for the government to run welfare to work programs. This evidence led, in California in particular, to increases in welfare to work programs when other budget items were being cut during the budget crisis of the mid-90s.

In the early 80s, low cost job search welfare-to-work programs delivered modest but positive earnings, but they didn't leave families better off. Families who went to work weren't making more than they were getting from welfare. This led to some programs experimenting and seeking to do better by these families, often through higher-cost education and training programs. But the research here showed that education and training focused programs were not significantly better than low cost job search programs. The basic consensus of these evaluations was that the best route was a mixed strategy. Riverside County in southern California applied a mixed strategy from 1995–2000 (the program was changed to reflect research findings) and it ultimately succeeded in raising its effectiveness without federal policy change.

Before Congress passed time limits on welfare, MDRC conducted three studies of time-limited welfare. These results were of great interest to the states, particularly absent federal opinion. If the studies had found that time limits hurt families, state reactions would have been very different—no impacts make a difference. Similarly, in the 90s, a national Job Training Partnership Act (JTPA) study found positive impacts on employment and earnings for adult men and women, but negative or no impacts on out-of-school youth. Shortly thereafter, resources for youth programs were cut.

As a final example, MDRC rigorously tested the impacts of providing earnings supplements to low-wage workers, some of whom were welfare recipients, and some of whom were not. There were reasonably consistent and positive impacts across a range of locations on employment, earnings, job retention, and child outcomes. Benefit-cost studies were also consistently positive. This led to states adopting ideas from this debate, which they would have never otherwise done.

Robert Slavin presented on his work at Johns Hopkins on education programs for impoverished schools, Title 1, which provides for evidence-based research that can be used to develop alternatives for programs and policy. An example program is Success for All (SFA), a comprehensive program for grade schools from preschool through grade five with a focus on reading. It's one of the few programs that exist that has been rigorously studied. There are over 50 studies that have looked at the impact of SFA. One national three-year longitudinal randomized study found an effect size of .28 on reading scores and a cost per student of $130 per year for three years.

Slavin is also involved in reviewing research on a wide variety of programs, in trying to influence policy to adopt evidence-based approaches, and in expanding the amount of funding available for research. Research-proven programs are those that use control groups, are randomized, amass data over a period of at least a year, have been taken to scale, and have been proven to work a number of times in a number of places. Congress has supported the use of research-proven programs in legislation a number of times, and there is a lot to be learned from these examples.

One of the first of these programs was the Comprehensive School Reform (CSR), introduced by Congressmen David Obey (D-Wis.) and John Porter (R-Ill.) in 1997. At that time, there were several examples of full school reform models, each quite different, but relying heavily on professional development and research-based strategies. Seventeen programs were mentioned in the legislation, and several others were coming along. The legislation provided grants of up to $500,000 per year over three years, but schools had to promise to use a proven and comprehensive school reform model. More than 2,500 schools were funded before the program was terminated. Experiences were mixed: Though it was easy to monitor comprehensiveness, it turned out to be much harder to measure effectiveness. You could widely disseminate school reform initiatives but it was hard to evaluate them. About half of

the funding went to schools that had included strategies that not only hadn't been proven, but hadn't even been piloted.

However, a small number of schools went to proven models, and they discovered it is possible to have a large-scale government program that includes an incentive to use proven strategies where schools would follow through on implementing these strategies, and by and large, would experience positive results. Even though the program has been gone for a while, there are hundreds, maybe thousands of schools using CSR, and these improvements have lasted. In the case of schools who adopted SFA, improvements have lasted an average of seven years, which is extraordinary for school reform. One important feature is that CSR wasn't telling schools what to do, but instead was allowing them to pick from a variety of models, so that they had choice within a set of high-quality options. If you look at schools overall, the results of CSR are not that impressive. But when schools adopted one of a variety of externally-developed, comprehensive, proven programs, then it has tended to work.

The second model is Reading First (RF), which is a program that is a part of No Child Left Behind (NCLB) and is currently in the process of being discontinued. Reading First was intended to be a showcase of NCLB, the prime example of applying evidence-based approaches to teaching kids to read in low achieving schools. Results were disastrous, in part because RF was based on theoretical research rather than what had been evaluated and found to be effective. Adopting this looser definition meant that consultants and government employees basically had an incentive to pick programs that they liked. When these programs were implemented and then subjected to rigorous research, they were found not to be effective.

From these two experiences, legislators have learned about what works and what does not. If you give competitive grants to schools to use models that provide strong evidence of effectiveness, and you tell them what makes this evidence strong, then some schools will come forward. It is also essential to give schools choices: the choice to not participate at all, and the choice of what program to use. If you do this, and have the follow-up to make sure that programs are actually implemented as intended, then there is the possibility to make a big difference on a substantial scale. There would also be side benefits because significantly more research and development would be funded by the government and by other parties. So, if you get the idea of proven effectiveness as an industry standard, you would get a cycle of reform as people are motivated to find and evaluate what works, and to create more high quality programs.

Jon Baron spoke about his organization, the Coalition for Evidence-Based Policy (CEBP), a nonprofit, non-partisan organization dedicated to promoting government policy based on rigorous evidence about what works. They have a bi-partisan board and work closely with OMB, Congress, and the Department of Education. They are not affiliated with any specific programs or program models, which enables them to remain neutral. The Coalition has helped advance some concrete reforms, such as the evidence-based nurse family partnership, which was included in the President's budget. They also worked with the Judiciary Committee to get funding for evaluating programs under the Second Chance Act.

One of the first examples of research impacting policy was welfare policy: Programs that focused on moving people into the workforce with short-term job searches or job training were more effective than programs that emphasized remedial reading, math, and basic education. There was strong evidence that the more effective programs were cost-effective, and the more effective programs turned out to save the government millions. This didn't require a full benefit-cost analysis, but just an analysis that showed that the costs of using the reformed welfare to work program was less than the costs of the

control group. This type of evidence may not have had a huge impact on Republicans, who were already persuaded, but it helped persuade the centrist Democrats to support a work-focused reform.

The nurse visitation program is still early in the process of impacting policy, but it is a leading example of a research-driven program. The program provides nurse visitation for women who are low-resource and pregnant. The visiting nurse comes during the pregnancy and during the first two years of the child's life, and provides information on parenting, nutrition, smoking and drinking cessation, and birth control. The program has been evaluated by three gold-standard randomized trials with long-term follow-up, and has been found to be highly effective. For example, the program results in 40–70 percent reductions in child abuse and neglect and lower rates of arrests and conviction of the children of those mothers, compared to the control group. There is also a significant impact on cognitive and educational outcomes for the child.

These studies led directly to the creation of the new Evidence-Based Home Visitation Program at the Department of Health and Human Services. The appropriations language for this program also mandated evidence-based programs throughout the states, and told states not to add additional elements that have not been subjected to rigorous testing.

A key ingredient in these examples that allowed for successful adoption and impact was that the research was randomized, controlled trials, conducted in real-world settings, showing sizable and lasting effects on outcomes. This type of research tends to be persuasive to policymakers, because non-academics can understand it. More complicated research designs, such as those using econometric designs to control for differences, tend to be less accessible, and therefore less persuasive to policymakers. With complicated studies, policymakers may wonder about hidden assumptions and people trying to game the system.

This implies that BCA researchers should try to keep research or benefit-cost methods as simple and accessible as possible so that they are understandable to non-researchers. Use randomized studies to estimate costs and benefits, and in other areas, such as estimating shadow prices. It will also be critical to report findings in clear, plain language with little or no research or statistical jargon. Jargon will exclude policymakers who would otherwise use the results. Finally, strive for simplicity in results, as well as in methods.

Q and A Session for Panel Five

QUESTIONER: Some policy can't be subject to random studies. Also, if we hold up random assignment studies as the gold standard, we risk undermining other research tools, and ignoring many things that may be helpful to us. We'll never have the resources to do random studies on everything that matters.

WALLACE: I agree that there are many areas that can't be investigated with random assignment trials. For example, it's very had to study, say, saturation initiatives. Given the costs, they have to be chosen very carefully, and they should be aimed at identifying what works (a particular strategy) rather than "does this program work?"

Findings in one arena were persuasive enough to start trying in other policy areas—the idea of financially rewarding certain behaviors. It has been applied in community colleges and in a public housing demonstration project. So this idea, that of financial incentives to reward good behavior, has had a broader impact. Another finding is that group-based interventions can be more effective—

interventions that target a cohesive group, rather than individuals, tend to consistently have better outcomes. There's a good theoretical basis for this kind of strategy working.

QUESTIONER: One of advantages of evaluation-based research is that you keep it simple, and this means that you aren't necessarily asking why it works. Then you are left with the question of whether you know why something works. In general, in more complex research designs, you may be trying to get at *why* something worked. And then you may get at something that suggests a general principle that you can apply to other policy area. Are the results transferable?

BARON: Most random assignment studies look not just at differences between the means, but also at the factors that led to the effect, so the results are more comprehensive than you would think. They are hypothetical, but important, and ultimately the causal factors can be tested in follow-up trial. For example, in the nurse studies I mentioned, the fact that it was nurses as opposed to paraprofessionals appeared to be an important factor. So a follow-up study looked at nurses, paraprofessionals, and no intervention, and nurses proved to be better.

ANTON: Sometimes it's not morally acceptable to have a control group. For example, if we want to follow-up on a person who is in a control for drug addiction treatment, and we want to follow up for 20 years, that's just not ethical. But the other point I would like to make is that if you can show success through evaluation, then you create an incentive for evaluation. Currently, we find that people do not want to be involved in evaluation, and in particular do not want to pay for evaluation, because they feel it will take money away from the program. We would get many more and better studies if evaluation were perceived as being in the interest of providers to have these studies, rather than something imposed by governments or other external forces.

SLAVIN: I think you raise two different questions. One is how do you get people excited about participating? And the short answer to this is that you pay them off. We often use delayed trials so that people who are in the control group get the treatment a year later. The other issue is how you get them excited about implementing? And I think this is also about money. For example, when you are talking about resources, poor schools are highly motivated by even a small amount of money. So if you have even a hint that evidence will matter, then people become very enthusiastic about it.

QUESTIONER: It seems that we often create programs because we have a problem, and not necessarily because we have a solution to that problem. Do you think it's possible to change that culture, so that it would be possible to attach a "study" requirement to funding in Congress?

ANSWER: The short answer to this is yes. We included $20 million to HHS so that they could take the most promising programs they had and subject them to rigorous study. We also put in funding and a requirement for the Census Bureau to conduct a longitudinal (seven year study) to follow a set of randomly assigned individuals who had had a confirmed case of neglect or abuse, and there was a book written about it. In a nutshell, this is Jon Baron's mission in life, which is to put this requirement up front in the legislation.

Missing Shadow Prices from Benefit-Cost Analyses of Social Programs

David Weimer: What shadow prices can we develop that will be useful across a variety of social programs, and is it possible to develop them in such a way that they will be credible, and therefore widely useable? There are many examples of shadow prices, for example, the value of a statistical life has been the subject of many studies, so many that there are multiple meta-studies of the value of a life in the United States. The range is pretty wide, but most economists would place the value between $2 million and $6 million. But there are many other less well-known shadow prices, for example, transportation analysts are confident that the value of commuting time is roughly 50% of wages.

Robert Haverman: Nearly all BCA studies of important social policies require a number of shadow values. This often requires that one turn to other sources to find these values to calculate the impact of a program. Consider BCA of the Federal Section 8 Program, benefits and costs of the program accrue to voucher recipients, non-recipients, including the government, and society as a whole.

[Corresponding image not currently available.] Components in red are directly estimated in the research program and include the earnings effects and the employment effects of the program, and the costs that are devoted directly to it. These items do not involve shadow values. For all the rest of the values, there are no easy-to-find values that they can attach to the impact, and it's on those items that a shadow value is necessary. For example, to the extent that a voucher recipient moves to an improved neighborhood, or a house that is of higher quality, there is going to be a willingness to pay for that benefit, but that won't be picked up, and one needs to be able to estimate the elasticity of the demand curve, or the shadow value, for that benefit. Another example of values that need to be accounted for is the increased years of schooling of children who are recipients of vouchers. Any full benefit-cost analysis is rife with the need for shadow values and prices.

For example, how do you provide a shadow price for the value of a child who moves to a better neighborhood and gets more schooling? Assume that through their study, one can estimate that the years of schooling gained are attributable to the program. Then, the question becomes how you value increased years of schooling. There are 14 items that describe one component of the benefits related to valuing additional schooling and only two of these items are private market returns: the estimates of increase in earnings attributable to schooling (and those values are in the education literature). An additional nine items are individual benefits that are not valued on the market, and three items are public and social returns, also not valued on the market. We will have to estimate the shadow values for *each* of these non-market benefits if we want to reliably estimate the positive impact of an extra year of schooling.

To consider just a few nonmarket gains from additional schooling:
- His or her spouse will also experience marginally increased productivity.
- The person's children will also have better health and more education as a result of their schooling.

Chair:
David Weimer is a Professor of Public Affairs and Political Science in the La Follette School of Public Affairs at the University of Wisconsin, Madison.

Panelists:
Robert Haveman is John Bascom Emeritus Professor in the La Follette School of Public Affairs and the Department of Economics, and a Faculty Affiliate at the Institute for Research on Poverty at the University of Wisconsin, Madison.

V. Kerry Smith is the W. P. Carey Professor of Economics at Arizona State University.

Philip Cook is ITT/Terry Sanford Professor of Public Policy Studies, Professor of Economics and Sociology, and Associate Director of the Terry Sanford Institute of Public Policy, at Duke University.

- The individual themselves will have better heath, higher life expectancy, will have a lower probably of disablements, will make a better marital match, and will make better consumer decisions.
- And looking externally, there will be reduced crime and greater community involvement.

However, despite the daunting nature of this task, there are methods available for securing shadow values, in particular private non-market values. This method exploits the relationship between schooling and marketed inputs, and is used to produce non-market outcomes. They would start with studies that establish a correlation between education and some non-marketed outcome. The improved health of an individual must have a reliable coefficient estimate relating schooling to the variable of interest as well as variables with market values that are likely to be associated with that outcome. So for example, physician's bills, the cost of policing in a community, private music lessons, income, are all variables that are likely to be related to some income, and these values have dollar amounts attached to them.

This equation is one that is in all introductory economics textbooks and certainly every intermediate micro textbook. It says that the ratio of marginal product to prices will be equalized across inputs in the production of any output. The price of schooling in this example is the price that we want to know—the implicit willingness to pay to produce a particular non-market output. So if we rearrange this equation, we now get the price of schooling (the shadow value we are interested in) on the left hand side. If the marginal products of schooling and the other input are equal, then one can tease out the shadow value on additional schooling related to that particular outcome.

As an example, consider an estimated equation that relates a mother's education and the dollar value of household income to the number of times a child repeats a grade. Using the estimated coefficients from that equation, an additional dollar of household income reduces the probability that a child will repeat a grade by 0.002. The mother's education is represented by a dummy variable that represents whether or not the mother has a high school diploma, and let's say that if the mother has a diploma, this reduces the probability a child will repeat a grade by 0.062. Using the formula, the marginal monetary value of the mother's high school diploma, on the probability that the daughter will repeat a grade, is $310. To get this result, we needed the 0.0002 value and the 0.062 value, and the fact that income is measured in dollars.

Here are a few estimates that are determined in this way:
- Cognitive achievement is up to $500 per year for an additional year of schooling.
- Gains in health equal a one-time payment of $9,000 for an additional year of schooling.
- Gains in the value of consumer decision-making equal $300 a year for an additional year of schooling.

We can attach dollar values (shadow values) to all the non-market items, and this has to be added to the dollar impact of the program, so that any analysis includes *both* the marketed and the non-marketed benefits. Our guess from estimating many of these shadow values is that the non-market value of these effects is *at least* as big as the marketed effect. And this means that the rate of return is approximately double what is commonly reported in the literature when you look at the effects of an additional year of schooling.

KERRY SMITH: Expanding on what was implicit in the previous discussion, there are two categories of effects: the private, or primary effects, which accrue to the individual who was the object of the program (who received more schooling, for example); and the public effects, which are external to the

person who received the schooling. But in defining these public benefits and assigning shadow prices to them, we bring up an important question of transferability to other situations, other areas, and other programs. In the process of defining those concepts, you also define an *objective* choice—exactly what is it that people feel they are getting as the result of a program. Also implicit in the prior discussions was a definition of the "extent of the market," which most people define as the user community, those that are directly affected by the programs. But in economic circles, there is another important group of non-users who are being left out here. These are the non-users who do not use the services, but are concerned about these areas. For example, people who want to save a wilderness area, even though they will never visit, and it may well be that there is a significant group of non-users who care about social programs as well.

How do we identify these people, and find out what their values might be? Economic value is about tradeoffs, and when people make choices, we call this a *revealed preference.* If we don't or can't observe people making choices, then *stated preferences* come into play. *Stated preference analysis* is about offering credible, understandable, and hypothetical choices for proposed financial consequences. This is a huge multi-billion dollar industry, though we hear very little about it, because it generates proprietary information.

How do you do it? The first thing that you have to do is to understand how people think about and describe goods (objects of choice). And then you have to define a mechanism, a plan that actually will begin to operationalize a change in that object of choice. It's not enough to say that you will magically remove families from poverty. This is where trials and databases would be helpful in defining the problem and outlining the proposed solution, and then you have to explain it in a way that people will understand. You have to evaluate information in this context to make sure people understand it. Once you've done that, you have to explain the choices that are available to them in a way that says, even if this is hypothetical, someone is going to take their "choice" seriously. This is not easy, but it can be done. If you can do it, then there is a very large body of research that says that you can develop reliable estimates of the values of the objects of choice through this process. How do you explain objects of choice? Perhaps the most famous example of this was the plan outlined for the Exxon Valdez spill, which used a series of actual pictures. These surveys cost about $300 per person, but the costs have declined dramatically with the advent of more sophisticated data technology. The researchers outlined choices that outlined options to help prevent a future spill, and the choices that they were presenting were real. One of the challenges was that they had to clearly explain the concept of risk to ordinary people.

Have contingency valuation surveys been done for social programs? Only two. One was a limited study in Gijon, Spain in which they were trying to determine what payments people would accept in lieu of home care services. They were already users of the program, so they were asking a willingness to accept question, and calculated an amount per hour that the people said they were willing to accept for each hour of care. Fifty-five percent of the people refused compensation, but this shouldn't be viewed as a failure; rather, it means that people didn't want to accept a trade, and that they were perfectly happy with home care. No one asked how the rest of society felt about this choice—the object here was defined as only the users. So this means that no one asked what the rest of society would be willing to accept in lieu of the services.

How should researchers start to look at this question for social programs? If we are really serious about this, then we should adopt a model that was successfully used to evaluate energy models some time ago. We need a shadow price modeling forum and two or three independent teams charged with

making their analysis transparent—providing every level of detail about what they do in the process of their research. They should be forced to work independently, and to meet periodically over three years with a group of external reviewers that includes both supporters of the efforts to evaluate programs, and skeptics. It wouldn't cost a lot of money, and the payoffs are potentially huge.

PHILLIP COOK: Crime is an outcome of concern in a variety of social programs, not only in criminal justice, but also educational programs, early childhood programs, job training, and Section VIII Housing. Each of these programs raises the question of how we value reductions in crime. There are two approaches to valuation. First is ex post valuation of the crime, which includes loss to the victim, lost productivity or the cost of injuries, cost to a third party, and criminal justice follow-up. This approach has been embraced particularly by the public health community in looking at the cost of disease.

Approach two is the ex ante valuation of safety from crime, or the value of feeling safer. This approach, which is endorsed by most economists, recognizes that a significant cost to community is the cost we incur in trying to avoid becoming victims. This plays out in decisions such as where we live, where we shop, where our kids go to school, where we go to dinner or park, and the overall economic development of an area. All of those costs are incurred before any crime is committed. Crime control has cost to taxpayers and to our individual freedoms, and even after all of the prevention, there is the value of the remaining risk of victimization.

In effect, crime is a disamenity, like pollution or trash. An example of this is assessing crime-related benefits of a lead-exposure reduction program because lead exposure correlates to increases in crime due to its health effects on kids. How do we think about the benefit in crime reduction? Should we follow the first approach, and put a value on the crime that didn't happen, or should we instead take an ex ante approach? If youth are less crime prone, then benefits include: lower likelihood of victimization; less private avoidance, which means people will go out more, or will feel safer sending their children out to play; lower criminal justice expenditures; lower expenditures on home protection devices; and less fear. Over the long term, there will be revitalization, more commercial activity, and enhanced economic development. That's the concept: Crime is a pervasive threat that we adapt to, so reducing it has economic benefit for everyone.

Willingness to pay for safer community has begun to be measured by contingent valuation surveys. For gun violence, we created a nationwide sample referendum that sought to find out how much each household would pay for a 30 percent reduction in gun violence. In fact, by randomizing our surveys, we were able to construct a demand curve for the reduction of violence, which averaged $270 per household for a 30% reduction in gun violence. And this includes all of the variety of benefits that households will get from a reduction in crime. Similar studies have been conducted by others, for other types of crime, ranging from a valuation of $25,000 for burglary, to $9.7 million for murder.

There are some problems with WTP estimates. One of the most obvious problems is that the questions specifically target a change in the crime rates, but that's not the same thing as safety. For example, if a "little old lady" hides in her apartment, her victimization is zero; if the neighborhood improves, and she feels safe enough to goes out, her rate of victimization may be higher, but her quality of life will also be better. It's also very hard to get a handle on the quantitative impact that reduced crime has on economic development—the change in property values is a good proxy, but it is difficult to isolate the effect of reduced crime in practice.

One really important issue that comes up in these kinds of crime studies is the question of standing. The question is whether the criminal should have standing or not? Should the offender "count" when it comes to valuing crime and punishment? If so, theft (particularly cash) is a transfer without cost, rather than a loss to the victim of the theft, but the cost of theft then becomes what we spend trying to avoid being victimized. Thus, vandalism has a higher cost than theft. The question of how we count benefits and costs is particularly tricky for crime issues.

What's next? Continue developing the WTP approach because it's better to be imprecisely right than precisely wrong and contingent valuation approaches are at least action in the right direction, even if they still need to be improved. We can learn from the analogy between crime and disease. Lastly, we need to confront the standing issue.

Q and A Session for Panel Six

QUESTIONER: How do you decide when it's better to estimate and when it's better to have no number at all?

HAVEMAN: Responding with a question, do you feel as comfortable with a sheet that just estimates the shadow price of an extra year of schooling? Or do you feel better if I say, I really don't know, but I'll give you my best crack at it, and here's how I'm going to break it down and think about it? The benefit of the latter is that it gives you something to debate and to talk about.

RESPONDENT: We really need to try to get at these numbers, but at the same time recognize and respect the uncertainty involved, and try to bracket our results and do sensitivity analysis to ask whether it really matters, when possible.

SMITH: We saw in Steve Aos' presentation that it's possible to take the literature and distill it in a careful, informative way. But one would never suggest that a single set of shadow values is the answer. Instead, the answer is going to be an accumulated set of values, which will allow us to develop an understanding of attractors and the factors that influence (or don't influence) the values. Secondly, the absence of information does not mean the decision gets shelved—the decision still has to be made. The real question here is this: are we going to try to inform those decisions, and understand that two or three benefit-cost analyses are not going to provide all the answers.

WEIMER: The scholarly rewards are for finding small, precise answers, even if it's a narrow piece of the puzzle. We need more comprehensive efforts to think broadly about policy areas, and we need to acknowledge the areas where we are not measuring well.

QUESTIONER: I appreciated Dr. Smith's analysis of non-use values, and I'd like to hear some thoughts on dealing with problems of non-paternalistic altruism. I'm thinking of public health where we have been focused on parent's views of children's health and where current approaches take parent's valuations of their children's health. How should we think about the source of their concerns?

SMITH: This is a long, protracted debate. When Exxon Valdez took place, there were a small number of tourists whose activities were prevented, a larger number of commercial fishermen were precluded from their work, and perhaps even larger numbers of indigenous communities who were affected in significant ways—a relatively small cohort. But if you say that there are a hundred million households in the lower 48 who care about this, even if they each care a small amount, that's a big number. What is

the source of those concerns? Is it because the people in the lower 48 care about the tourists, fishermen, and Native Americans, and how happy they are? If so, then this is okay under some rubrics. Or is it because they care about the assets?

The fact of the matter is that in stated preference models, we are trying to explain the connections that people are making. And there are at least as many connections that suggest that people are acting because of welfare economics so as to suggest that they care about the outcomes and the wellbeing of the people, and the practical answer is that we will never know. So I prefer to adopt the reasons that say that they prefer the social outcomes, which is perfectly consistent with the outcomes we see.

QUESTIONER: When you use the ex ante approach to crime, you get estimates that are generally at least 3 times the estimates you find using the post ante model. This raises the question of whether we are inflating the estimates. My other question is how did you link the increased probability of being a victim of a crime with a plausible program, and can you tie that probability to some specific type of outcome?

COOK: You're right and the ex ante approach should be higher because it takes into account effects that had been ignored under the other approach. The values are higher, mostly because it does a better job of accounting for all the costs. Can we learn from looking at people moving to opportunities? One of the top reasons housing residents gave for wanting to participate in activities was to reduced crime for themselves and for their children. And in fact, we see lower blood pressure and better health as a result. It would have been interesting to ask them to place a value on this, but we didn't.

HAVEMAN: The approach that I took was clearly an ex-post approach, which assumes that markets are functioning. I could also use Phil Cook and Kerry [Smith]'s strategy of asking people about their willingness to pay for the non-market effects of education. My guess is that, consistent with the results that we were just talking about, the results from this approach would be higher than the one that I showed you. What troubles me is, at least with the ex-post strategy that I was using, people feel pretty comfortable that this is grounded in a sense of markets and the way they work. I don't have a very good feel for how I would use a CV or ex ante approach to valuing the non-market effects of schooling. It's an interesting question to think about, and I must say I'm puzzled by it.

QUESTIONER: Dr. Smith alluded to benefit transfer in environmental regards. How does that research today translate into social policy?

SMITH: First of all, you have to begin by recognizing that we will never know the true values, and that every valuation of a benefit transfer process involves some set of assumptions. Let me give you an example. If you evaluate a program that is operating in location 1, and then you try to use the results of evaluation 1 to impute what the results would have been in location 2, that second program is different—the people are different, the location is different, and likely even the program itself is different.

A better tool is to use meta-analyses. The results of a meta-study will not tell us what the truth is, but the ability to use a statistical approach leaves us with less unexplainable variation between studies, and improves our confidence.

STEGMAN: I just wanted to mention a couple of things. First, we just funded a national stated preference survey on willingness of American taxpayers to commit to reduction in child poverty and it is in its early stages. There has also been a small stated preference survey in 12 or 13 states on willingness

to pay to move towards a rehabilitation oriented model for juvenile justice in place of the model that is currently common within the mainstream justice system.

You'll see in Jonathan's talk today that none of us are benefit-cost experts. The genesis of MacArthur's interest is really a political philosophy. We believe that dealing with the problems of poor people really generates benefits for all of us. So this is a way of bringing the country together, and not isolating people, or cutting social programs. Jonathan [Fanton] will talk about evolution of his thinking in this regard, but the notion of non-users is very important. Sometimes we think of cost-avoidance, but this doesn't really value how we feel about having our children grow up in a society that is unjust. Hopefully, some of the stated preference and willingness to pay will help us look at non-users who are less likely to be benefiting directly, but who still have an interest in seeing these efforts be supported. This notion of benefits to participants and to society is important conceptually to us. How we do that methodologically is still a question we have. Should we do a stated preference survey that estimates cost of reducing child poverty by one child? I don't know the answer to this question—we will have to come back to it as we think about the architecture of this larger effort that we are involved in, and prioritize, because we can't fund everything, even if it is important—we just don't have the resources.

We're hoping to stimulate other activity by other funders so that there are multiple studies on the same issue and we can begin to see the variation—many of these non-monetized impacts could be very long term.

Panel 7

Chair:
Scott Farrow is a Professor and Chair of the Department of Economics, University of Maryland, Baltimore County.

Panelists:
Arnold Harberger is a Distinguished Professor of Economics at University of California, Los Angeles, and the Current and Founding President of the Society for Benefit-Cost Analysis.

Lynn Karoly is a Senior Economist with the RAND Corporation.

Lester Lave is a Professor of Economics at Carnegie Mellon University.

David Weimer is a Professor of Public Affairs and Political Science in the La Follette School of Public Affairs at the University of Wisconsin, Madison.

Issues in the Development of Principles and Standards for Conducting Social Benefit Cost Analysis

Arnold Harberger: In each year's budget, the government says what it plans to spend, and plans to receive, and it never works out that way. So there is always a margin of adjustment between budget and reality. How does this work? Ninety-nine percent of the time it's in the capital markets, which act like a sponge that day by day really works for government, and that is a good justification for arguing for capital market sources.

Of course, if you do use capital market sourcing, and someone says I have a project that will be funded through an increase in the gas tax, then we make one module where we finance the project from the capital market, and a separate module that collects the money from the gasoline tax and then dumps it into the capital markets. And then we merge these two modules to get the result. So the adoption of a consensus concerning capital market sourcing is our standard way of doing things. This doesn't preclude non-market funding.

If we have a capital market source, we ought also to incorporate a shadow price for government funds. The general idea is if you have a project like a public electricity enterprise that pays own way in cash, then you don't need to do any adjusting in the benefit-cost analysis.

But if you have a road that is not a toll road whose benefits are social (time savings, cost reductions to users), then over time there will be the initial debt expense, plus the interest payments. Since there's no money coming in to pay the interest, the debt keeps piling up through the life of the project. If we don't put extra cost on it, then our explicit assumption is that we will keep rolling it over forever, even after the death of the project. Incorporating a shadow price is fairly easy, and direct, and therefore, it should be done.

With respect to the discount rate, one can get a discount rate that reflects using the capital market sources. But people have used, and to some degree continue to use, a variety of other discount rates. If one uses another discount rate, one is logically forced to use a shadow price of borrowed funds. This recognizes the fact that by borrowing, you are displacing a lot of investment that might have occurred, and you have to recognize this through the shadow price.

We can reach a consensus that we have to avoid multiple discount rates. In other words, we cannot value a park, which is going to give its benefits in kind, through users. We cannot take their present value willingness to pay, because among the users, their discount rates will vary. Because of this, we have to value all benefits and costs in the moment in which they are occurring on the cash flow basis. This is a problem that is not widely recognized. Once you think about it, it makes sense.

A final plea: We all are motivated by considerations of meeting basic needs, fighting poverty, paying special attention to different disadvantaged groups, *and* distributional weighting is for the birds. But basic needs externalities and premiums for benefits accruing to special groups—these strategies are

much more viable. The problem with using distributional weighting is that if they are any distance apart (such as 0.5 and two), then it will allow the user to justify taking money away from someone, and giving it to someone poorer, even if you are wasting a huge majority of the money (in the example above, up to 75 percent). Meanwhile, if you make the distributional weights very close (such as 1.01 and .99), then you don't get any mileage.

Lynn Karoly: One goal of BCA is to help us make decisions about choosing between programs within a particular domain of social policy, such as within early education. We are also interested in making decisions across these domains—for example, helping public policymakers decide whether to invest in early childhood versus within grade schools. We are often asked if it is better to do preschool programs or to reduce class sizes in grades one through three.

Even if we might want to compare across domains, we often haven't taken the tools to a level that allows us to make these comparisons fairly. Within early childhood programs, there is an extensive set of rigorous experimental evaluations that demonstrate that well-designed programs can be effective, but only a subset have had BCAs done as part of their evaluations. Even within that subset of programs, BCAs are not done in the same way and this makes it hard to compare among different interventions.

Out of 20 rigorously evaluated early childhood programs, only seven have measured benefits and costs. When we look at benefits and costs overall, there is a very wide range overall, which is contrary to the principle we heard yesterday that you should be looking for the one biggest number. What these results give us is proof of several general principles. But these principles don't take us all the way to truly comparable numbers that can be used to make decisions about in which programs to invest.

The general principles:
1. It's clear that not all programs are exceeding costs. In some cases, this is because there were no effects. But in other cases it is because the effects were all in things that are hard to put a monetary value on (and the amount of time that you follow-up has a big impact on this).
2. Favorable effects are found both in smaller demonstration programs and in large, publicly funded programs.
3. You can have a positive ratio from higher cost, more intensive programs, but also from lower cost, less expensive programs.
4. Returns can be higher with targeting. You get bigger benefits by targeting to needier populations.

However, it's still difficult to compare among different interventions, pointing out the potential role that standards and principles will play.

There is one set of issues within principles and standards that will be easier to deal with: discount rates, age discount, making decisions about which stakeholders to include or whether you are looking at total benefits for society, accounting for uncertainty, and other issues. In the programs that we looked at, some of these were already dealt with, and then we could make adjustments for others so that they were more comparable.

The harder issues to address have to do with which outcomes are measured, the extent to which you follow up with participants and track longer-term outcomes, whether you project outcomes beyond the follow-up period into the future and who you project them for, and what shadow prices you attach to these outcomes. For example, each of the programs we looked at was evaluated by a different team of

researchers, and each team tended to measure the set of outcomes they thought were most relevant. But they may have benefits in other areas that were unmeasured, and other programs may have measured those benefits. The Visiting Nurse program focused on mothers, but also measured benefits for children; whereas, many of the center-based programs measured benefits for children, but didn't measure whether there were collateral benefits for parents.

Follow-up is another difficult issue. The Perry Preschool program followed participants up until age 40, but many other programs ended follow-up relatively soon after the end of the program. You can project benefits into the future, but one lesson of the Perry Preschool program is that projections of future outcomes can actually understate benefits—at each successive stage of follow-up, the benefit-cost ratio has increased, as they have actually been able to observe the benefits, rather than project them. So the upshot is that you are on unequal footing in measuring outcomes.

Another issue that comes up for these programs is for whom to measure benefits: participants, families, peers, or schools? Should neighborhoods and benefits to descendents also be included in the measurements? These studies vary in who they study and how they measure, so the challenge is to find a standard.

Shadow prices are another issue. This is particularly pertinent to social programs because there is a huge array of benefits that require shadow pricing. For example, we looked at 39 evaluations of social programs. Of those programs, 22 had BCA, but looking at domains of impact, they included outcomes for children and youth, and sometimes even children and youth of the parents. Likewise, there are set of adult outcomes, or they may be outcomes measured for participants when they reach adulthood, and within each of those domains, there is actually a whole array of outcomes that may be measured.

There are more challenges in these areas in terms of bringing the approach to standardization. It's important to do so if we want to be able to use these tools to make decisions, but it won't be an easy process.

Lester Lave: Some of the basis for BCA is utilitarian theory—the "greatest good for the greatest number." When we are using BCA, we have to pay attention to the distributional effects and explicitly recognize the utilitarian basis of the analysis. There is also a need to recognize that increasing efficiency can lead to "one size fits all." If you have private markets, then people can choose what they would like to do; whereas, if you have a regulation, then you don't get that kind of choice.

You also have to look at valuations. When you are doing BCA, you are acting as if there were uniform values, and that's difficult. For example, if you say that the value of a life is worth $6 million, some act as if they value their lives at $6,000 or even $600 (and it's unclear whether we can have any impact on peoples' valuation of their own lives). Context is an enormously important piece of this. For example, a salmon swimming in the Yukon River has enormous value, potentially tens of millions of dollars; whereas, salmon on a bed of ice at your local fishmonger is going to be approximately $10 a pound. I would offer that example as a caution to the question about what ability we have to generalize. Context is *really important* to almost everything, and this really complicates the benefit-cost analysis, and particularly the ability to generalize.

[In the mid-1960s, Lave was part of a project to calculate the benefits and costs of conducting benefit-cost analyses at the Army Corps of Engineers. His introduction to this topic was through speaking with someone who had been at the Corps a long time who told him a story about how Lyndon Johnson had

asked the Corps to conduct an analysis of making the Trinity River (Texas) navigable all the way upstream, and the Corps said it didn't make any sense. Then Johnson asked them to do a BCA analysis, and they did the analysis, and came back to him and said that the benefits outweighed the costs.]

Recently there was a story about a proposal to build levees along the Mississippi River so that we wouldn't have another tragedy like Katrina. If you have a levee 50 feet high all the way down to the Gulf, all that water has to go someplace and it just wouldn't work. Before, the idea was that you built levees around the towns to protect them and then the floodwaters would spread out over farmland. So every 15 years or so, it floods the farmland—and now the farmers are saying that they would like to be protected as well. This idea of building levees all along the Mississippi is now being seriously discussed.

The benefit-cost analysis isn't complete, but you can already tell that the benefit-cost analysis is more about good sense—the result of building walls all the way up is that you are going to flood St. Louis. What you see here is that some of the benefits are being calculated out to four significant figures. There are really no cases in which you have enough confidence in your calculations that you should have four significant figures, and even if you need two significant figures to make the benefits outweigh the costs, this might really be a tie.

This is a bothersome issue. In all work on air pollution abatement, we get overwhelming benefits from reducing premature mortality, and almost no benefits from improving visibility and smell. It makes you think that there's something wrong with benefit-cost analysis because it doesn't match with common sense. Here, you see that people are usually giving conservative values—we want analyses to be based on best estimates, not conservative values—because conservative values are arbitrarily biased in one direction, which means that we'll make bad decisions part of the time.

Another issue has to do with depletion of environmental ozone. If there is a depletion of ozone, there will be more UV, more tanning, and more skin cancer. So they calculated what would happen if more Americans got skin cancer, and immediately, the benefit-cost calculation changes so you are compelled to deal with ozone. But then we realized that make-up has sun block in it, which then suggests that you should encourage women to wear make-up, which then changes the benefit-cost calculation completely. This is another case where we need to let common sense, rather than the benefit-cost calculation, guide us.

We also looked at a benefit-cost analysis of a NASA program for the Pentagon. They wanted to know whether it was worth it for them to be paying for 10 percent of the research and design costs for supersonic transport (Boeing was paying the other 90 percent). We did our analysis, and showed that Boeing would make back about 10 cents on the dollar of their research costs. So of course, Boeing immediately asked the government to start paying 90 percent of the research costs. NASA immediately concluded that they wanted us to repeat the benefit-cost analysis; we did, and found that the benefits were 40 cents on the dollar. We went down to Washington, and finally were left alone with one man, who said to us, "You have embarrassed me twice in front of my boss—you tell me what benefits I need to assume so that benefits outweigh the costs."

And the question this raises is who within government is doing benefit-cost analysis? Although there are a few celebrated examples of benefit-cost analyses that are done well, in general, they are not done well, and in some cases they're done by someone like the man in the previous example. The message here is that we can work to improve the theory, but there are many times in practice where it will still

be implemented poorly, and there's nothing that theoreticians can do about that. We should be modest about where BCA can make a difference. If we expect too much, the tool will fall apart.

David Weimer: Who would set the standards? One extreme would be to have a standard-setting body, and a public rule-making process that determines how benefit-cost analyses should be done if they are paid for by federal dollars. At the other extreme would be a decentralized process, which is what we have now, which means that academics would review the literature in a particular area, assess practice, and suggest ways for making it better. An intermediate approach would be something like the standard setting in industry, where process is managed by consensus with industries involved. In our case, you can imagine a relevant professional organization taking on this role. But our Society for Benefit-Cost Analysis is a little to small and a little too young to play that role, and many larger professional societies don't want to take this on—so we're left with the question of who will do it.

The second relevant question is what would the standards cover? You can look at this along one of two axes. On the first axis, the content of the standards could look at methods of analysis (as we have been talking about so far), or standards of transparency, which is where we should put our efforts. On the other axis, you also have a question of how binding the standards are, from advisory to mandatory.

In the northwest quadrant we have advisory standards that address methods. Examples of these include the UK Treasury Green Book and EPA guidelines for preparing economic analysis. These are both very sophisticated and fairly comprehensive, but advisory. This is what we should do, but not what we have to do. The NOAA panel on contingent valuation is another example of advisory standards, but it is a one time shot, which asks whether contingent valuation is ready for use, and if so, what methods should be used. These have been positive efforts. I don't know of any systematic study of how influential these have been, but I know contingent valuations often cite these standards, and they have been cited thousands of times.

Moving to the southwest, here's where we have mandatory rules about the content of the analysis—required techniques and specified parameter values—like the old OMB Circular A-94. A-94 required that everyone use a seven percent discount rate, which is too high. The point, though, is not particularly that it was too high, but that it stayed in place too long because it was mandatory. This characterization is supported by the literature comparing federal regulations with private consensus processes, which found that public standard-setting efforts are generally episodic, very infrequent, and making relatively major changes. The private standard-setting efforts, on the other hand, are characterized by many small incremental changes.

Now, let's go to the northeast quadrant, where we have advisory standards that give advice relating to transparency. I have a hard time finding examples of this that are from benefit-cost analysis, but one example is in the social sciences. Many social science journals require that an author archive his data in a public place, so that peers can review it. If the Society forms a journal, I would suggest that we require that authors keep an electronic archive with all the details of their study so that anyone who wants to do a similar analysis can go back and review it, find enough detail to know how it was done, and decide how they might adapt that method to their needs.

Turning to mandatory transparency, an example of this was an e-government mandate that required regulatory agencies to put dockets on the internet. This is transparent and democratic—because anyone who wants to can see them, and can add a comment. We could argue that anytime an agency completes a draft BCA or receives a deliverable from a contractor to do BCA, they would have to post it

and give an opportunity for comments. This would really strengthen public involvement in the process, and would increase accountability and transparency. Also, if it was a salient issue, losers and winners would be able to make their case in support of costs and benefits that they thought were missing from the analysis.

One final caution to having methods-based standards: If NOAA had carried out their contingent valuation study 10 years earlier, they probably would have concluded that the tool wasn't ready to be used. And if this had been published, it would have impeded development of the tool, with the result that we would be 10 or 20 years behind where we are today in our ability to use it. The danger of judging too soon is high for methods-based standards—high enough that we might be better off sticking to requiring more transparency.

Q and A Session from Panel Seven

QUESTIONER: I've worked on projects in real estate where there are public and private benefits and we've used different discount rates for the different flows of benefits. So how do you discount the social benefits?

HARBEGER: You've pointed out exactly the problem. When a business firm does discounting, there's nothing wrong with a private firm using 20 percent cost of capital, and when we do a job for them, we should use their rate. But when we use varying rates for social benefits, from the standpoint of society as a whole, we can get in trouble. Until the profession deals more directly with this problem, social analyses should have a common rate.

RESPONDENT: Using a different discount rate within the same analysis to reflect differing risk is wrong. Instead, you should use a scenario probabilistic distribution model to take account of the risk.

QUESTIONER: I don't think that guidelines and sunlight are mutually exclusive. If you have both, people can respond to standards. Net benefit is also an important measure. Is it ratios or net benefit? Instead, you should be using the opportunity cost of capital for your discount rate, which may vary.

QUESTIONER: This is a comment for Dr. Weimer. I don't think these things are mutually exclusive. I think that we absolutely need standards, but that's not mutually exclusive with the effort to shed a little more sunlight.

Secondly, I heard a lot of people talk about benefit-cost ratios. OMB, on the other hand, says that they prefer to discuss net benefits. To my mind net benefits might be better, but in any case, I would propose that this is one topic that standards should address.

QUESTIONER: I want to make one point, and then ask Dr. Harberger a question. In this discussion of principles and standards, only the older people in the room will remember that there were two sets of principles and standards that came out of the old water resources agency, including a very comprehensive set from the Water Resources Council. In my view, both of those had impacts. Then there was a third one that came out of GAO. My question to you Al [Harberger], is if you could clarify why you are opposed to distributional weightings—if we could get the citizenry to accept them, then why not use them?

HARBERGER: If I say that your rate is one-half, and Dick Zerbe's rate is two, the gap of 75 percent that I talked about is there, and it's an invitation for me to go out looking for projects which still waste money. I'm pushed to love a project that wastes half of the money, and really love a project that only wastes a quarter. The point is that someone will infer from the tax result alone, and this is a problem if you apply it rigorously to all possible and actual decisions.

Now when you get to something like dealing with poverty, you can say that you attribute a positive externality of raising poor peoples' income from $10,000 to $12,000 per year, so we'll give them an extra 50 percent on top of the financial impact—that is something that is easy to deal with. So there are times that this makes sense, when you're talking about specific projects. But I don't agree with producing these distributional weights across all projects.

DISCUSANT: If there's another mechanism by which to re-distribute wealth, and the other way of doing it is less costly than the project, then you can't give the re-distributive effects any more weight in the project.

DISCUSANT: Guidance exists on this point from the OMB, which says that you treat all dollars as equal. And then if you're concerned about distributional effects, you do a separate analysis. And this is basically the end of the question.

HARBERGER: Basic needs externalities is a different way of approaching distributional weights, and I want to emphasize that here. What I don't agree with is people who have a penchant for using distributional weights across the board. People like to use them where they get nice elegant examples, but they don't see the areas where the logic of that system forces them into absurd results that no one would want to accept.

RESPONDENT: I want to make a connection. The externalities you are talking about are the same things that we were talking about as having been measured often in non-real purpose ways, and that's one reason I think there's great potential to include these.

Click here to read "Philanthropy, Benefit-Cost Analysis, and Public Policy," Remarks by Jonathan F. Fanton, president of the MacArthur Foundation, at the 2008 Benefit-Cost Analysis Conference.