

Testimony of James Sullivan
Rev. Thomas J. McDonagh, C.S.C., Associate Professor of Economics,
University of Notre Dame
Director, Wilson Sheehan Lab for Economic Opportunities

Before the U.S. House of Representatives
Committee on Agriculture
Subcommittee on Nutrition

July 15, 2015

I. Introduction

Chairwoman Walorski, ranking member McGovern, and other members of the subcommittee, thank you for inviting me to participate in today's hearing, "Past, Present, and Future of SNAP: Developing and Using Evidence-Based Solutions." I am talking to you today because examining evidence on the impact of programs designed to help the marginalized and disadvantaged has been the focus of much of my academic research over the past 15 years.

Recently I co-founded the Wilson Sheehan Lab for Economic Opportunities (LEO), which is a research center at the University of Notre Dame that identifies innovative, effective and scalable programs that help those in need move to self-sufficiency. The Wilson Sheehan Lab has partnered with some of the largest private providers of services to the poor in the country, including the Catholic Charities network, as well as state and local governmental agencies, to examine evidence of program effectiveness. These projects include programs to promote community college completion, comprehensive case-management programs, homeless prevention services, and diversion programs for first-time, non-violent offenders. This evidence allows resources to be channeled to the programs that will have the greatest impact on the lives of the most vulnerable.

As I have learned through my work at the Wilson Sheehan Lab, the idea of impact evaluation is a foreign concept to many local service providers. While these front-line providers are driven by compassion and motivation for helping the poor, most of them design and launch programs without solid evidence of effectiveness. They typically measure outputs or track outcomes for program participants, but rarely are the programs evaluated using rigorous methods. The lack of evidence of what works and what doesn't is not unique to local programs. The same could be said of many national programs that operate on a much larger scale.

II. The Need for Greater Evidence on the Impact of Social Programs

The New Era of Experiments

One of the greatest advances in the social sciences in recent decades is the development and application of methods that allow us to determine whether social programs are having their intended effect. The general idea behind these methods is simple: one can determine the impact of a program by comparing outcomes for a group of people who are exposed to an intervention (the treatment group) to those for a group who are not (the control group). The control group is supposed to reflect the counterfactual—what the outcomes for the treatment group would be if they had not been exposed to the intervention. The gold standard approach is the randomized controlled trial (RCT), in which individuals are assigned to either the treatment or control group randomly. In cases where an RCT is not feasible, quasi-experimental approaches are available that are intended to mimic the research design of an RCT. For example, one might compare those eligible for a program to those just barely ineligible; or compare program participants to those on a waiting list. These alternative approaches typically do a much better job of determining the impact of interventions than nonscientific approaches such as tracking outcomes for program participants. However, if the control group does not appropriately reflect the counterfactual, the quasi-experimental results are not reliable.

Nowadays, gathering experimental evidence is commonplace in many sectors. The medical profession runs tens of thousands of experiments each year to test the effectiveness of new pharmaceuticals, medical procedures, devices, or treatment regimens.¹ These experiments have led to vast improvements in healthcare all across the globe. Large companies are constantly running experiments in an attempt to better target resources, attract new customers, or avoid spending money on projects that don't work. Google runs 20,000 experiments each year while Capital One runs three times that many.² Shouldn't the same commitment to proven effectiveness apply to our social programs?

The Impact of Evidence

By steering resources towards the most effective social programs, evidence of what works and what doesn't can significantly improve the lives of the poor. All too often innovative, promising interventions are not brought to scale because program operators are unable to demonstrate effectiveness. Greater evidence of impact for these successful programs would attract the resources necessary to serve more disadvantaged individuals and families. For example, several RCTs have shown that the Nurse-Family Partnership, a home visitation program for new, mostly low-income mothers, has improved outcomes for both mothers and children.³ This evidence has helped spur a national home visiting initiative.⁴

¹ Manzi, J. (2012). *Uncontrolled: The Surprising Payoff of Trial-and-Error for Business, Politics, and Society*. New York: Basic Books.

² Prigg, M. (2012, February 6). "The human search engine." *London Evening Standard*. Retrieved from <http://www.standard.co.uk/news/techandgadgets/the-human-search-engine-7315344.html>; Brooks, D. (2012, April 26). "Is Our Adults Learning?" *The New York Times*. Retrieved from http://www.nytimes.com/2012/04/27/opinion/brooks-is-our-adults-learning.html?_r=0.

³ <http://toptierevidence.org/programs-reviewed/interventions-for-children-age-0-6/nurse-family-partnership>.

On the other hand, ineffective programs often persist because there is no evidence to show they don't work. These programs squander precious resources that could have greater impact elsewhere. Or worse, they may cause harm to those they intend to help. One notable example is Scared Straight—a program that aimed to deter juvenile delinquency by exposing at-risk youth to prison life and adult inmates. The program gained national attention when it was featured in the 1978 Academy Award winning documentary, *Scared Straight!* Based on anecdotal evidence of success from several studies that did not meet minimal scientific standards, many states and other countries adopted the program. When the program was eventually evaluated using experimental methods, several studies showed that the program actually “led to higher rates of offending behavior.”⁵ A 2004 Washington State Institute for Public Policy cost-benefit report concluded that one dollar spent on a Scared Straight program resulted in an additional \$264 in costs in today's dollars.

By guiding funds away from ineffective programs, high quality impact evaluations allow us to do more good with the limited resources available. This not only produces better results for those in need, but also for the economy as a whole.

Evaluating Social Programs

There are many examples where large-scale experiments have informed social policy. In the early 1990s more than 30 experiments were run at the state level to test the impact of changes in features of welfare programs. The results of these experiments—most noticeably the effect of work incentives on employment—helped shape landmark national welfare reform legislation in 1996.⁶ Other notable experiments include the Head Start Impact Study and the U.S. Department of Housing and Urban Development's Moving to Opportunity demonstration project. Experiments such as these are becoming much more common.

However, when Google alone runs thousands more experiments than the government agencies that oversee social programs, it is clear we have not come far enough. And the fact of the matter is that funding decisions are typically made without information on program effectiveness. Only about 1 percent of non-defense discretionary dollars are backed by any hard evidence.⁷ Most domestic social programs are not evaluated, or are evaluated with unreliable methods. For example, in 2009 federal agencies spent about \$18 billion on 47 employment and training programs, but as a recent GAO report concluded: “little is known about the effectiveness of most programs.”⁸

⁴ See Chapter 2 of Haskins, R., & Margolis, G. (2014). *Show Me the Evidence: Obama's Fight for Rigor and Results in Social Policy*. Brookings Institute Press.

⁵ Petrosino, Andrew, Carolyn Turpin-Petrosino, Meghan E Hollis-Peel, and Julia G Lavenberg (2013). “‘Scared Straight’ and other juvenile awareness programs for preventing juvenile delinquency.” *The Cochrane Collaboration*.

⁶ Grogger, J., & Karoly, L. A. (2005). *Welfare Reform: Effects of a Decade of Change*. Cambridge: Harvard University Press.

⁷ Nussle, Jim and Peter Orszag (2014). “Let's Play Moneyball,” in *Moneyball for Government*. J. Nussle, & P. Orszag, eds. Disruption Books.

⁸ U.S. Government Accountability Office, “Providing Information on Colocating Services and Consolidating Administrative Structures Could Promote Efficiencies,” GAO-11-92: Published: Jan 13, 2011. <http://www.gao.gov/products/GAO-11-92>. Discussed in Barnes, Melody and John Bridgeland (2014). “Making Moneyball Work,” in *Moneyball for Government*. J. Nussle, & P. Orszag, Eds.. Disruption Books.

Evaluating SNAP

Although SNAP is the nation's largest cash or near cash means-tested transfer program (with costs exceeding \$70 billion annually in recent years⁹), there is little hard evidence on the impact of the program. While no large-scale experiments have evaluated SNAP,¹⁰ there is some promising quasi-experimental evidence from recent research showing that in utero exposure to the food stamp program is associated with increased birthweight and lower rates of obesity and heart disease in adulthood.¹¹ But this evidence is for those exposed to food stamps in the late 1960s and early 1970s. There is a clear need for rigorous, experimental evidence of the impact of SNAP in its current form. Is SNAP the best way to improve nutritional outcomes for the disadvantaged? The honest answer is that we don't know. Bringing evidence to bear on this question would allow us to better help those struggling to put food on the table.

An important challenge to evaluating SNAP is that the structure of the program can make it difficult to implement the most rigorous experimental or quasi-experimental methods. It is much easier to conduct experiments when a program is rolled-out, expanded, or changed in significant ways, or when the program is not made available to all those who might be eligible. Food stamps is an entitlement program that has been around for more than five decades, and although there have been notable changes to the program, these reforms are modest compared to those for other programs such as AFDC/TANF or the Earned Income Tax Credit.¹² Also, because states are given limited flexibility to experiment with changes to program rules it is difficult to gather evidence on the effectiveness of possible improvements to SNAP.

The most recent Farm Bill made some important strides towards encouraging more policy based on evidence by authorizing \$200 million to support 10 pilot projects designed and implemented by state agencies to reduce dependency and encourage work.¹³ Each of these pilots is required to have an independent evaluation that compares outcomes for households participating in the pilot to a "control group" of households not participating in the pilot. The legislation also requires the participating states to make administrative data available in order to track outcomes. It is important for pilot programs such as these to require or incentivize grantees to evaluate their pilots using the best experimental or quasi-experimental methods possible in order to generate the kind of evidence needed to shape future food assistance policy.

⁹ <http://www.fns.usda.gov/pd/supplemental-nutrition-assistance-program-snap>.

¹⁰ See <http://www.fns.usda.gov/ops/supplemental-nutrition-assistance-program-snap-research> for a list of recent SNAP research, which includes some small-scale RCT studies.

¹¹ Almond, D., H. Hoynes, and D. W. Schanzenbach. 2011. "Inside the war on poverty: The impact of food stamps on birth outcomes." *Review of Economics and Statistics* 93(2): 387–404. Almond, D., H. Hoynes, and D. W. Schanzenbach. 2012. "Childhood Exposure to the Food Stamp Program: Long-run Health and Economic Outcomes," NBER Working Paper No. 18535.

¹² For further discussion see Hoynes, H. W., & Schanzenbach, D. W. (2015). U.S. Food and Nutrition Programs. NBER working paper 21057.

¹³ <http://www.fns.usda.gov/2014-snap-e-t-pilots>.

III. Promoting Evidence-Based Policy

I applaud this subcommittee's efforts to seek ways to generate more evidence to guide the future of nutrition policy. There are a number of strategies to encourage the development of a stronger base of evidence and to ensure that this evidence is used to target resources towards programs that work and away from those that don't. Let me highlight a few:

1. Incentivize innovation

First, you can't expect programs to be built on evidence of effectiveness if there is no evidence. The most innovative ideas for social programs frequently come from states or local providers. But state and local agencies and private service providers often lack the resources to put these ideas into practice. Thus, we need funds to encourage providers to experiment with new, promising ways to help those in need, and to build strong evidence for innovative programs. A nice example of this approach is the Department of Education's Investing in Innovation (i3) initiative, which has distributed more than \$1 billion in grants using a tiered-evidence model to fund programs to improve student achievement. The lowest tier, or "Development," i3 grants provide support for promising initiatives that currently lack rigorous evidence. Grants such as these create a pipeline of innovative programs that, if proven effective, can be scaled up to ensure broad impact.

2. Incentivize programs to be based on rigorous evidence

Second, new funding should go predominantly to programs that can provide solid evidence of effectiveness. In this tiered-evidence approach, funds are allocated by merit-based competitions, as opposed to formula grants where geography or other factors are more important than rigorous evidence. These competitions create a market for proven solutions. Here, again, the i3 initiative offers an excellent framework. The evidence requirement for the top tier, or "Scale-up," i3 grants includes one or more well-designed and implemented RCTs or quasi-experimental studies.¹⁴ Such a tiered-evidence model could be incorporated into future SNAP pilot programs, similar to the one authorized in Section 4022 of the most recent Farm Bill.

3. Require initiatives to be rigorously and continuously evaluated

Third, even when new programs are grounded in solid evidence, it is important to ensure that they are rigorously evaluated. Otherwise, there is no way to know whether the program is being implemented correctly and having its intended effect. Many government grants that support social programs require grantees to evaluate the program being funded. A typical evaluation might only track outcomes for program participants. In the case of a SNAP pilot to promote work, this might mean that the grantee provides information on earnings and employment for program participants. While this information is valuable, it is not sufficient for determining a program's impact, because we don't know the counterfactual—what the outcomes for these participants would have been had they not participated in the program. Better information about program impacts would be available if evaluations were required to be well-designed and well-implemented experiments or quasi-experiments.

¹⁴ For further discussion of the tiered-evidence model see Haskins and Margolis (2015).

A program should continue to be evaluated as it is scaled-up, even when there is clear initial evidence of success from an RCT. Program impacts can sometimes be difficult to replicate in other settings. For example, experiments testing the impact of re-employment bonuses in Illinois showed strong evidence of success, but subsequent studies of similar programs in other states were less promising.¹⁵ Positive results can be hard to replicate because the expanded program may not be implemented correctly, or the success of the initial program may hinge on a feature of the intervention that is hard to replicate, such as an extremely talented program operator.

Evidence of program impact is also helpful when the results are less promising. Often times, modest or negative results can uncover issues with program design or implementation. In this way, evidence promotes an environment of learning that leads to better programs.¹⁶ But if evaluations of a program continue to produce disappointing results, then the funds for this program should be reallocated towards efforts with evidence of success. There tends to be a fair amount of inertia in social policy—funding today goes to the same programs that were funded in the past. Even when hard evidence shows a program is ineffective, it is often difficult to pull the plug. Consider the case of the Even Start Literacy Program, a national initiative established in 1989 that was designed to improve both child and parent literacy. Three national evaluations showed that the program had little impact—children and parents in the treatment group “did not gain more than children and parents in the control group.”¹⁷ Even after the release of these findings more than \$1 billion were allocated to the program, and it was more than ten years before the program was finally defunded.¹⁸ In order to best leverage evidence to improve outcomes for children and families, there needs to be willingness to shut down ineffective programs.

4. Provide greater access to administrative data

Fourth, one of the most significant barriers to high quality impact evaluations is limited access to administrative data. Collecting data on participants in an evaluation can be an expensive proposition. An hour long survey can cost upwards of \$500 per completed survey. In many instances, administrative records already collect information on key outcomes such as employment, earnings, program participation, college enrollment, income, criminal history, and many others. Moving to Opportunity and other large scale impact evaluations have relied heavily on these kinds of administrative data. The problem is that these data are often not available for evaluation purposes. Some cities and states have established administrative data repositories that can be used for evaluation, but there needs to be a national effort. One promising example is the National Student Clearinghouse that compiles student enrollment and degree information for

¹⁵ Meyer, Bruce D. (1995). “Lessons from the U.S. Unemployment Insurance Experiments,” *Journal of Economic Literature*, Vol. 33, No. 1. (Mar., 1995), pp. 91-131.

¹⁶ Sperling, Gene (2014). “A Continuum Approach,” in *Moneyball for Government*. J. Nussle, & P. Orszag, eds.. Disruption Books.

¹⁷ U. S. Department of Education, Planning and Evaluation Service, Elementary and Secondary Education Division, “Third National Even Start Evaluation: Program Impacts and Implications for Improvement,” Washington, D.C., 20202. <http://www2.ed.gov/rschstat/eval/disadv/evenstartthird/toc.html>.

¹⁸ Bridgeland, John and Peter Orszag (2013), “Can Government Play Moneyball?” *The Atlantic*, July/August. <http://www.theatlantic.com/magazine/archive/2013/07/can-government-play-moneyball/309389/>.

more than 3,600 colleges and universities across the country. The data from this clearinghouse provide researchers, educators, and policymakers with the information necessary to advance evidence-based education policies.

In conversations with Chairman Ryan’s staff prior to the introduction of the Ryan-Murray commission bill, we discussed a document I wrote about how a clearinghouse for program and survey data might work.¹⁹ This clearinghouse could provide policymakers and researchers access to administrative data on dozens of government programs. These data would be accessible, on a restricted basis, through a centralized but secure information system that would allow users to link participants across programs, to respondents in surveys, and to other administrative data sources. These data would provide critical information on program impacts that is currently unavailable to policymakers and researchers. This would make possible countless studies of government programs, and as such, would transform the way researchers analyze and evaluate these programs, and provide policymakers with better evidence of program impact and effectiveness, resulting in the design of more effective government programs.

The Ryan-Murray Evidence-Based Policymaking Commission Act of 2015 represents an excellent step towards greater access to data to evaluate programs. This legislation would create a commission to study how administrative data on federal programs and other data might be compiled in order to facilitate research and evaluation. The commission would also study the feasibility of a national clearinghouse for such information.

5. Disseminate evidence on what works

Finally, policymakers, educators, service providers, and researchers need a way to track down easily the existing body of evidence on what works and what does not. A national repository of well-designed, well-implemented impact evaluations would help promote a broader culture of evaluation. An important challenge here is that the body of evidence on social programs is far larger than the body of good evidence. Stakeholders need a way to filter out unreliable studies to ensure that policy decisions are guided by the most reliable evidence. This means that we need clear standards for what constitutes solid evidence. Ideally, an independent entity would assess evaluations and identify those that are reliable.

A well-designed model of how to synthesize a large body of evidence is the What Works Clearinghouse, which is run by the U.S. Department of Education’s research arm: the Institute of Education Sciences (IES). For this clearinghouse, IES reviews research on education programs and policies, and highlights those that are of the highest quality. The U.S. Department of Labor offers a similar service for labor topics through the Clearinghouse for Labor Evaluation and Research (CLEAR). Outside the government, the Coalition for Evidence-Based Policy provides a nice one stop shop for what works in social policy. It has established the highest standards for identifying social programs that are supported by well conducted RCTs.²⁰

¹⁹ Sullivan, James X. (Forthcoming), “Promoting Greater Understanding of the Impact of Federal Programs: What a Clearinghouse for Program and Survey Data Might Look Like,” Wilson Sheehan Lab for Economic Opportunities Policy Brief.

²⁰ <http://evidencebasedprograms.org/>.

IV. Conclusions

Advances in technology and data collection have greatly expanded opportunities to implement high-quality evaluations of social programs. Evidence from these evaluations can help in the design of programs that yield better results and guide policy on how best to allocate scarce resources. By encouraging innovation and evaluation and by targeting support at interventions shown to be successful, policymakers will ensure that our social programs are more effective at helping vulnerable populations get ahead. We at the Wilson Sheehan Lab welcome this transformation in the way we fight poverty in America.