

BUYING OUTCOMES

LESSONS FROM THE PAST

/ **PAUL RYAN** FORMER SPEAKER OF THE HOUSE

An earlier version of this chapter appeared in the first volume of Workforce Realigned in 2021.

America is at a moment of great need and great opportunity in the fight against poverty. With so many events and innovations shaping our economy, the importance of disrupting the stale institutions in place to tackle these challenges has become clearer than ever before.

I've been concerned for years by our country's approach to reigniting upward mobility. Our efforts have too often originated in Washington, D.C., with little input from those with lived experience and from the individuals working on the ground to expand opportunity. This effort has led to approaches that further displace and marginalize those living in poverty. It's when we innovate together to solve this problem — combining the vibrancy of community-based solutions, the know-how of the private sector, and the scale of government policy — that we have the greatest potential to make a difference.

Pay-for-success programs bring together the best of the public and private sectors to address the most critical issues our country is facing. The goal of a social public good — a world in which far fewer Americans live in poverty — is central to their execution. So too is the expertise and capital of the private sector, which can provide funding, strategic thinking, and energy to deploy resources where most needed. When executed properly, programs like these offer enormous promise.

Unfortunately, the long history of performance-based contracting in American civic life includes frequent examples of programs that have not achieved their desired results. Over the past four decades, the government has attempted to structure several programs that offer payouts to impact investors based on provider performance to drive better outcomes. Although these programs were created with the best intentions, they have driven little improvement to the status quo. Identifying and addressing the challenges they have faced will be critical to designing the next generation of performance-based contracts.

For example, a common shortcoming of performance-based contracts is that many fail to differentiate payouts according to the level of need of the target populations. Without adjusting payments to account for different levels of risk and vulnerability among participants, service providers are penalized for serving people with greater needs. Initiatives across workforce development, education, and health care have made this same mistake, creating incentives that work against delivering services to the populations most in need of support.

When performance-based contracts are set up well, shortfalls are mitigated, and the programs have significant potential to improve lives. Well-executed performance-based contracting offers benefits for all parties involved by shifting spending risk away from governments, creating positive feedback loops based on provider effectiveness, and facilitating the collection of data on intervention outcomes.

▶ ...a common shortcoming of performance-based contracts is that many fail to differentiate payouts according to the level of need of the target populations. Without adjusting payments to account for different levels of risk and vulnerability among participants, service providers are penalized for serving people with greater needs.

As the next generation of performance-based contracts takes hold, it is essential that we learn from past challenges. Historical examples from workforce development and health care offer lessons on how to mitigate typical shortcomings and fully unlock the potential of performance-based contracting.

JOB TRAINING PARTNERSHIP ACT (1982)

Going back to the 1980s, federal legislation has tied payments to employment outcomes achieved by program participants. The '60s and '70s saw the rise of several federal training programs but few that led to positive results.¹⁹⁴

/ FIGURE 1 /
KEY MILESTONES OF PERFORMANCE-BASED CONTRACTING



Unemployment rates continued to increase among groups targeted by government workforce development funding, and even large-scale programs like Job Corps failed to show significant sustained impacts.¹⁹⁵

In an effort to support higher-quality job training initiatives, the federal government, through multiple programs, began allocating dollars to local jurisdictions partially based on participants' employment outcomes. The Job Training Partnership Act (JTPA) of 1982 was one such program. Developed through a bipartisan effort led by former Sens.

194 James Bovard, "The Failure of Federal Job Training," Cato Institute, Policy Analysis no. 77, August 28, 1986, www.cato.org/policy-analysis/failure-federal-job-training.

195 U.S. Department of Labor, Office of Inspector General — Office of Audit, "Report to the Employment and Training Administration: Job Corps Could Not Demonstrate Beneficial Job Training Outcomes," March 30, 2018, www.oig.dol.gov/public/reports/oa/2018/04-18-001-03-370.pdf.

Dan Quayle, Edward Kennedy, Paula Hawkins, and Claiborne Pell and by former Reps. Augustus Hawkins and James Jeffords, it was signed into law by then-President Reagan. The bill aimed to improve employment rates for low-income Americans by providing budgetary rewards and sanctions to jurisdictions based on the near-term labor market outcome levels achieved by participants.

▶ Well-executed performance-based contracting offers benefits for all parties involved by shifting spending risk away from governments, creating positive feedback loops based on provider effectiveness, and facilitating the collection of data on intervention outcomes.

To carry out its purpose, the JTPA established federal assistance for adult and youth programs, federally administered programs (such as training for migrant workers and veterans), summer youth employment and training programs, and training assistance for workers affected by layoffs. The program established a performance management system that provided rankings of 620 service delivery areas (SDAs) and set aside funding to reward SDAs that performed particularly well relative to the overall labor market.

To evaluate outcomes, the JTPA originally considered four performance measures: rate of entering employment, average wage at placement, cost per participant who entered employment, and rate of entering employment among welfare recipients. However, states were given considerable flexibility to select comparison data and define favorable terms. Where improved results existed, it became clear that they had been driven by the selection of participants who had fewer needs and, therefore, were easier to serve.¹⁹⁶

By the early 1990s, the JTPA was spending \$1.5 billion annually in federal and state funds to provide employment and training services.¹⁹⁷

However, a 1991 report from the U.S. Government Accountability Office revealed discrepancies in the services offered to women and minorities, affirming the limitations to the program's results and reach.¹⁹⁸

196 Patrick Lester, "The Promise and Peril of an 'Outcomes Mindset,'" *Stanford Social Innovation Review*, January 13, 2016, www.ssir.org/articles/entry/the_promise_and_peril_of_an_outcomes_mindset.

197 U.S. Government Accountability Office, "JOBS and JTPA: Tracking Spending, Outcomes, and Program Performance," July 26, 1994, <https://www.gao.gov/products/hehs-94-177>.

198 U.S. Government Accountability Office, "Job Training Partnership Act: Racial and Gender Disparities in Services," July 17, 1991, <https://www.gao.gov/products/t-hrd-91-42>.

The JTPA suffered from several common challenges, as identified in an analysis by Burt Barnow and Jeffrey Smith:¹⁹⁹

- **The program provided stronger incentives to serve less vulnerable populations:** JTPA incentives treated all program participants equally, which led to higher margins for service providers who chose to serve lower-need individuals. The program did not serve groups such as women and people of color in proportion to their share of the eligible population, while individuals who would likely have achieved high post-training earnings regardless of the quality of the training were disproportionately represented. An outcomes-measurement structure assigning different levels of value based on the specific population's needs might have addressed this challenge.
- **The timing of performance incentives skewed the services provided:** In some cases, the length of the training programs was influenced by program managers' desire to count participants in their data for a particular program year. These arbitrary timing changes were found to reduce the overall mean impact of the training services the program provided. Updates to monitoring and reporting systems may limit the extent to which programs are able to manipulate data in this way.
- **Supportive evidence that incentives improved individual performance was limited:** It is unclear that the project improved the individual efficiency of employees in the absence of incentives at the individual employee level. Future performance-based contracts may explore how service providers can pass on incentive payments to their employees and how they can track individual performance without adding significant overhead costs.
- **Some providers gamed the compensation system:** There is strong evidence that JTPA service providers developed strategies to earn higher payments by gaming the performance system. A common gaming strategy involved formally enrolling participants in the program only after they had found jobs and then quickly terminating them from the program to increase the reported proportion of employed individuals. Adjusting reporting requirements and improving metrics and evaluation systems could help reduce the extent to which gaming can yield higher payments.

As a result of its structural challenges and the limited improvement to participants' employment outcomes, the JTPA was repealed in 1998. The program's failure to segment target populations, its focus on measurements that were not linked to individual performance, and its

199 Burt S. Barnow and Jeffrey A. Smith, "Performance Management of U.S. Job Training Programs," in *Job Training Policy in the United States*, ed. Christopher J. O'Leary, Robert A. Straits, and Stephen A. Wandner (Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 2004), 21–56, <https://doi.org/10.17848/9781417549993.ch2>.

lack of safeguards to avoid gaming the system are valuable reminders of the potential risks of performance-based contracting in workforce development. However, these mistakes also offer lessons regarding critical areas of focus for other pay-for-performance programs to succeed in the future.

TICKET TO WORK (1999)

In 1999, another performance-based workforce development program emerged that aimed to increase the number of low-income Americans achieving economic self-sufficiency. At the time, only 0.5% of Social Security Disability beneficiaries were leaving the benefit rolls because they secured jobs. Legislators hoped to create a better market to meet the diverse return-to-work service needs of beneficiaries, with the goal of doubling the rate of those exiting the program for work opportunities.²⁰⁰

The Ticket to Work and Work Incentives Improvement Act of 1999 was designed to support this mission by promoting flexible, customizable services to help disability insurance beneficiaries secure self-supporting jobs. The program incentivized private organizations and state agencies to deliver quality services by providing large payments for each client who secured a job and retained it for a long enough time to stop receiving Social Security Disability benefits. Beyond these financial incentives, the act was designed to support creativity and personalization by allowing organizations and beneficiaries to customize the programming available to each participant.

The Ticket to Work (TTW) program mailed “Tickets” to Social Security Disability recipients, which they could bring to a participating employer network to negotiate a set of services. For an employer network to receive payments, the participant was required to submit salary documentation. The program tied payment to long-term performance by requiring a beneficiary to stop receiving Social Security Disability benefits due to increased earnings for 60 months before the provider could earn full payment.

Despite the program’s intent to reach a broad group of beneficiaries, its early success was limited: By 2005, only 2% of individuals who received Tickets in the mail had used them, and only 45% of the 1,300 enrolled employer networks had accepted a Ticket.²⁰¹

Like the JTPA, TTW’s outcomes suffered from a range of shortfalls:

200 Craig Thornton et al., “Evaluation of the Ticket to Work Program: Initial Evaluation Report,” Mathematica Policy Research, Inc., February 2004, www.ssa.gov/disabilityresearch/ttw/ttw_report.pdf.

201 Craig Thornton et al., “Evaluation of the Ticket to Work Program: Assessment of Post-Rollout Implementation and Early Impacts, Volume 1,” Mathematica Policy Research, Inc., May 2007, www.ssa.gov/disabilityresearch/ttw3/ttw_report3.pdf.

- **Providers perceived the system as too financially risky:** TTW tied 100% of provider compensation to outcomes, which caused significant uncertainty as to whether payouts would be achieved. Research showed that after the first two years of program operations, employer networks relying on TTW payments as their sole source of revenue would have lost money: The cost-of-service delivery far exceeded TTW revenues for most providers. Offering up-front operating funding to providers in addition to subsequent outcomes-based payments, as many pay-for-success contracts now do, might have helped to mitigate this challenge.
- **The program provided higher payouts to providers serving less vulnerable populations:** Like other unsuccessful performance-based contracts, TTW created selection bias against harder-to-serve individuals and services less likely to lead to quick employment. Providers could refuse to serve individuals they thought were unlikely to maintain high enough earnings to stop receiving benefits and, therefore, unlikely to trigger outcome payments. They could also choose to offer services that aligned only with the outcomes payments they were likely to receive. Differentiating payment amounts based on participants' level of need could have helped avoid rewarding providers for serving clients with the lowest needs.
- **The benefits structure discouraged some beneficiaries from returning to work:** The program did nothing to address the "cliff" faced by participants who would lose 100% of their Social Security Disability benefits once their monthly earnings exceeded a certain threshold, which created a significant barrier for returning to work. Structuring the program with a sliding scale that reduced benefits more gradually might have increased the value proposition of returning to the workforce for participants.
- **Reporting and administrative burdens fell on service providers and participants:** Finally, significant administrative challenges delayed outcomes measurement and provider repayment. Beneficiaries were expected to submit salary documentation to employer networks but were given no incentive to do so, which made it difficult for employer networks to demonstrate that monthly earnings had reached the designated threshold. Establishing data-sharing provisions up front might have minimized administrative burdens and streamlined the system for triggering repayment.

From 2004 to 2007, TTW experienced a gradual decline in terms of provider interest and the number of Tickets assigned.²⁰² During that time, the program collected feedback from service providers, which suggested that TTW shorten the length of the payment period and offer larger payments earlier in the period. Providers also requested that the program award payouts for partial success and simplify the requirements for documenting participant salary levels.

In response to provider feedback, a set of revisions passed in 2008 that increased the number and total value of provider payments, shortened the period of participant employment for employer networks to receive full payment from 60 to 36 months, and revamped payment procedures to reduce the administrative burden. The revised system was significantly more attractive to providers, and the number of employer networks that accepted at least one Ticket doubled from 2007 to 2010.²⁰³

The increase in participation in TTW following the 2008 legislation reform affirms the importance of seeking service provider input to mitigate unforeseen barriers to entry. Things have certainly improved for TTW, which now benefits hundreds of thousands of Americans each year, but perverse incentives continue to be a challenge. Lessons from the program have demonstrated that getting performance-based contracts right takes consistent management and flexibility.

PHYSICIAN PAY FOR PERFORMANCE (2005–2018)

Health care is another area in which performance-based contracting offers both the potential to improve service quality and the risk of gaming and poorly structured incentives. The U.S. spends more on doctors, pharmaceuticals, and health administration as a percentage of gross domestic product than any other high-income nation, yet Americans do not enjoy better health outcomes.²⁰⁴ To combat rising costs and improve quality, states, health care systems, insurance companies, and federal agencies have piloted pay-for-performance (PFP) programs for physicians and hospitals across the country. The success of these programs, however, has been largely uneven.

In 2018, researchers from the University of Pittsburgh and Harvard University published a study reporting that Medicare PFP programs

202 Jody Schimmel et al., “Participant and Provider Outcomes Since the Inception of Ticket to Work and the Effects of the 2008 Regulatory Changes: Final Report,” Mathematica Policy Research: Center for Studying Disability Policy, July 25, 2013, [ssa.gov/disabilityresearch/documents/TTWNSTW%20Report-Final-072513.pdf](https://www.mathematicapolicy.org/disabilityresearch/documents/TTWNSTW%20Report-Final-072513.pdf).

203 Schimmel et al., “Participant and Provider Outcomes.”

204 Irene Papanicolas, Liana R. Woskie, and Ashish K. Jha, “Health Care Spending in the United States and Other High-Income Countries,” The Commonwealth Fund, March 13, 2018, www.commonwealthfund.org/publications/journal-article/2018/mar/healthcare-spending-united-states-and-other-high-income.

failed to improve health care quality or reduce costs.²⁰⁵ Rather than promote better outcomes, the program penalized physicians who cared for lower-income and sicker patients because the doctors' "quality scores," and therefore payments, decreased. The program's structure emphasized health outputs over baseline improvement, creating financial disincentives for doctors to treat patients who were less healthy.

▶ As we've seen, performance-based contracts can fall victim to predictable design errors. But when structured well, these programs have the potential for impressive results.

Providing higher payouts to those who serve healthier patients is a key concern about physician PFP programs, in which financial incentives often fail to promote health improvements over specific health outputs. While physician skill is an important component of health quality, factors such as the patient's baseline health, socioeconomic status, access to insurance, and exercise habits all contribute to health outcomes and are largely outside of the doctor's control. PFP programs that exclusively target physician pay without supporting other interventions draw a direct link from individual clinician skill to patient health that can create financial disincentives to treat the sickest patients.

▶ Health care is another area in which performance-based contracting offers both the potential to improve service quality and the risk of gaming and poorly structured incentives.

In one example of how a poorly designed PFP program created perverse incentives, a rural Veterans Affairs (VA) hospital turned away an ill 81-year-old veteran, even though the medical staff asserted that the veteran was too sick to return home. The hospital denied admission to the veteran with the understanding that, by limiting the number of sick patients admitted, the hospital would produce fewer bad outcomes. Rather than paying for success, the program's failed measurement standards enabled the hospital to improve its quality rating, receive greater funding, and earn a bonus payment for the hospital's director without driving real improvement in veterans' health outcomes.²⁰⁶

The lack of success of physician and hospital PFP programs has led many critics to call for an end to PFP in health care. However, it's

205 Kip Sullivan and Stephen Soumerai, "Pay for Performance: A Dangerous Health Policy Fad That Won't Die," *STAT*, January 30, 2018, www.statnews.com/2018/01/30/pay-for-performance-doctors-hospitals.

206 Dave Philipps, "At Veterans Hospital in Oregon, a Push for Better Ratings Puts Patients at Risk, Doctors Say," *New York Times*, January 1, 2018, www.nytimes.com/2018/01/01/us/at-veterans-hospital-in-oregon-a-push-for-better-ratings-patients-at-risk-doctors-say.html.

possible that an outcomes-based funding system could be effective in the absence of poor project design, weak measurement, incorrect outcome criteria, and flawed linkages between the intervention and outcomes. Past PFP programs struggled because they were structured around the underlying concept that financial rewards to physicians could improve outcomes in a vacuum. A stronger design could rescue the core concept.

FOUNDATION FOR EVIDENCE-BASED POLICYMAKING ACT (2017–2019) AND THE SOCIAL IMPACT PARTNERSHIPS TO PAY FOR RESULTS ACT (2018)

FOUNDATIONS FOR EVIDENCE-BASED POLICYMAKING ACT

A number of years back — before the COVID pandemic — we were getting frustrated in Congress about the lack of evidence in policymaking. At that time, we were stuck in ideological arguments on how best to fix problems and fight poverty through programs that were already implemented because we had very little data to prove what was working and what was not. We saw the positive impact in cases where we actually had measurements, and we were committed to bringing this approach to more programs in the future.

▶ Past PFP programs struggled because they were structured around the underlying concept that financial rewards to physicians could improve outcomes in a vacuum. A stronger design could rescue the core concept.

In 2017, Sen. Patty Murray and I introduced the Foundations for Evidence-Based Policymaking Act (“the Evidence Act”), which was intended to improve the ability of researchers, evaluators, and statisticians both inside and outside government to securely use the data that the government already collects to better inform important policy decisions. The Evidence Act, signed into law in January 2019, included 10 of the recommendations offered to Congress by the bipartisan U.S. Commission on Evidence-Based Policymaking, which was tasked with drafting a strategy for ensuring that rigorous evidence is created efficiently and as a routine part of government operations that could, in turn, be used to construct effective public policy. Key recommendations included improving data access, strengthening privacy protections, and enhancing the government’s capacity for evidence building.

SOCIAL IMPACT PARTNERSHIPS TO PAY FOR RESULTS ACT

As we've seen, performance-based contracts can fall victim to predictable design errors. But when structured well, these programs have the potential for impressive results. In support of improving the effectiveness of social services, Congress passed the Social Impact Partnerships to Pay for Results Act (SIPPRa), which was signed into law in 2018.

SIPPRa offers great promise for the next generation of performance-based contracts. The act appropriates \$100 million to the U.S. Department of the Treasury, \$15 million of which is set aside for evaluation costs to support state and local governments in building a foundation for outcomes-based decision making. Following initial delays related to the COVID-19 pandemic, in 2021 the Treasury Department announced more than \$19 million in funding for four jurisdictions to implement the first SIPPRa projects to address challenges from homelessness to job training²⁰⁷ and create opportunities to address the country's most pressing needs. The most recent round of SIPPRa funding is focused on projects that directly benefit children, with nearly \$41 million available for project awards and \$6 million for independent evaluations.²⁰⁸

▶ This evidence-based approach to lifting Americans out of poverty is directing funding to flow to programs whose methods have been evaluated using data, supporting real-world efforts that achieve positive results.

I'm personally incredibly proud of the Evidence Act, SIPPRa, and the principles reinforced by both of these complementary laws. This evidence-based approach to lifting Americans out of poverty is directing funding to flow to programs whose methods have been evaluated using data, supporting real-world efforts that achieve positive results.

In building programs based on evidence of what works, SIPPRa has the potential to finance the most effective solutions for fighting poverty, which originate not from Washington, D.C., but from leaders on the ground in communities across the country. SIPPRa funding will continue to support individuals and organizations that have been making a difference in their communities for decades while bringing their ideas to policymakers to expand their reach. This intersection of community-

207 U.S. Department of the Treasury, "SIPPRa — Pay for Results," <https://home.treasury.gov/services/treasury-financial-assistance/other-major-programs/sippra-pay-for-results/sippra-awards>.

208 U.S. Department of the Treasury, Office of Economic Policy, "Social Impact Partnerships Pay for Results Act: Differences Between FY19 NOFA and FY24 NOFA," <https://home.treasury.gov/system/files/226/FY19-NOFA-vs-FY24-NOFA.pdf>.

based approaches and government support is what will ultimately most improve the lives of Americans in need.

▶ In building programs based on evidence of what works, SIPPPRA has the potential to finance the most effective solutions for fighting poverty, which originate not from Washington, D.C., but from leaders on the ground in communities across the country.

SIPPPRA's evaluation requirement represents progress in the federal government's use of evaluations. However, too often the issue is not a lack of data collection — it's figuring out how to effectively use the data already collected. Detailed guidance on implementation of the Evidence Act, issued by both the Trump and Biden administrations, has helped identify and standardize best practices.²⁰⁹ Recommendations from the 2022 final report of the Advisory Committee on Data for Evidence Building (ACDEB) — created as part of the Evidence Act — are helping the Office of Management and Budget and other federal agencies share data in ways that address or ameliorate appropriate privacy and security concerns.²¹⁰ The National Secure Data Service, a bold new pilot project reauthorized by the bipartisan CHIPS and Science Act (2022), has a mandate to strengthen data linkage and data access infrastructure in service of evidence-based decision making across government.²¹¹

With a better federal system of evaluation — like the one mandated in SIPPPRA — we can identify programs that work. But we also need to have a better understanding of how to properly scale programs that have proven effective. This effort will require partnerships that go beyond government — partnerships with philanthropies, employers, community organizations, and more.

▶ It's time we carried forward these lessons to implement solutions that work for our communities, by educating policymakers, trusting the power of evidence, and incorporating the ideas of Americans who have confronted barriers to upward mobility for too long and deserve a voice in this fight.

With a new generation of performance-based contracts on the horizon, I'm hopeful we can learn from the challenges of past programs and continue to harness their momentum. It's time we carried forward

209 "Program Evaluation Standards — OMB M-20-12, Phase 4 Implementation of the Foundations for Evidence Based Policymaking Act of 2018: Program Evaluation Standards and Practices," Evaluation.gov, <https://www.evaluation.gov/assets/resources/Program-Evaluation-Standards.pdf>.

210 ACDEB, "Advisory Committee on Data for Evidence Building: Year 2 Report," October 14, 2022, <https://www.bea.gov/system/files/2022-10/acdeb-year-2-report.pdf>.

211 National Center for Science and Engineering Statistics, "The National Secure Data Service Demonstration," <https://ncses.nsf.gov/initiatives/national-secure-data-service-demo>.

these lessons to implement solutions that work for our communities, by educating policymakers, trusting the power of evidence, and incorporating the ideas of Americans who have confronted barriers to upward mobility for too long and deserve a voice in this fight.

Paul Ryan is the President of the American Idea Foundation and served as the 54th Speaker of the U.S. House of Representatives. In office from October 2015 to January 2019, he was the youngest speaker in nearly 150 years.