What We Know About the Impacts of Workforce Investment Programs

Burt S. Barnow
Johns Hopkins University

Jeffrey A. Smith
University of Michigan

November 13, 2007

This paper was prepared for the conference "Strategies for Improving [the] Economic Mobility of Workers" in Chicago, November 15-16, 2007, co-sponsored by the Federal Reserve Bank of Chicago and the W.E. Upjohn Institute for Employment Research. Neither co-sponsor nor, indeed, anyone but the two authors bears responsibility for the contents of the paper. Comments are most welcome.

Introduction

This paper briefly reviews the recent literature that seeks to evaluate employment and training programs, as well as important older papers. We focus on the bottom line question of whether or not the programs have measurable and economically relevant impacts on labor market outcomes.

We do not focus on the economics of such programs but do lean on the dismal science when interpreting the findings in the literature. We also do not focus on the econometrics of program evaluation, though our views about the credibility of various combinations of econometric strategies and data affects our choice of which evaluations to highlight and how we interpret the overall literature.

Readers interested in more in-depth surveys of the substantive literature should consult Heckman, LaLonde and Smith (1999). Smith (2000, 2004) provide a relatively non-technical guide to the evaluation literature, while Friedlander, Greenberg and Robins (1997), Heckman, LaLonde and Smith (1999), Angrist and Krueger (1999), Heckman and Vytlacil (2007) and Imbens and Wooldridge (2007) provide technical overviews.

Evaluations of the Major U.S. Federal Programs

Employment and training programs in the United States have a relatively brief history. In addition to the public employment programs of the Great Depression, the Manpower Development and Training Act (MDTA) [1962-172], the Comprehensive Employment and Training Act (CETA) [1973-1982], the Job Training Partnership Act (JTPA) [1982-1998] and the Workforce Investment Act (WIA) [1998-present] have provided vocational training, along with remedial education, subsidized on-the-job training and job search

assistance to disadvantaged youth and adults as well as displaced workers. CETA also provided public service employment.

Perry et al. (1975) review the literature on MDTA. Except for Ashenfelter (1978), this literature largely reflects the nascent stage of evaluation methodology at the time. The U.S. Department of Labor (DOL) funded a number of evaluations of the CETA program, all of which relied on the same data source, the CETA Longitudinal Manpower Survey (CLMS), which combined random samples of participants with non-experimental comparison group data from the Current Population Survey and included matched calendar year social security earnings data for both groups. Barnow (1987) summarizes these non-experimental evaluations, which relied largely on crude matching estimators and the bivariate normal model applied without exclusion restrictions and found widely differing impact estimates. Despite the high quality administrative outcome data, the CLMS lacked the detailed information on local labor markets found to be important in Heckman, et al. (1998) and the detailed information on recent labor market and program participation choices (at a fine level of temporal detail) found important in Card and Sullivan (1998), Heckman, et al. (1998) and Dolton et al. (2006).

The wide variety of CETA estimates led to a decision by DOL to evaluate JTPA using a social experiment, called the National JTPA Study (NJS), which operated at a non-random sample of 16 (of about 600) JTPA sites from around November 1987 to September 1989. NAME (19XX) describe the details of the experiment and NAME (19XX) present the results. The NJS included disadvantaged adults and out-of-school youth but not in-school youth and dislocated workers.

U.S. General Accounting Office (1996) provides impact estimates for five years

after random assignment based on Social Security earnings data. They find stable impacts of around $800 per year for adult (22 and older) men and women which lose statistical significance over time. In contrast, the estimates for male and female youth remain resolutely near zero throughout the follow-up period. As documented in e.g. Heckman et al. (2000), the NJS featured substantial treatment group dropout (FILL IN) and control group substitution into alternative providers of similar services (FILL IN). As a result, these estimates represent what Imbens and Angrist (1994) call local average treatment effects: average impacts on those who receive services if assigned to the treatment group who would not have received services if assigned to the control group. HLS (1999, Table 20) show that JTPA produced a set social benefit for adults but not for youth, pretty much irrespective of (reasonable) assumptions about benefit duration beyond five years, the discount rate or the welfare cost of taxation.

Mueser, Troske, and Gorislavsky (2007) use modern matching methods combined with relatively rich administrative data to estimate the earnings impact of JTPA in Missouri for program years 1994 and 1995 using a comparison group of individuals registering with the Employment Service. In real terms, their preferred estimates resemble those from the NJS.

Finally, although the WIA program has been operating nationwide since July 2000, there exist no published econometric evaluations. In November 2007, the Department of Labor announced a random assignment evaluation of WIA.

Evaluations of Other U.S. Programs

Job Corps

The Job Corps, established in 1964, provides intensive and comprehensive services, including vocational and academic activities as well as support services, to about 60,000 disadvantaged youth ages 16-24 in 119 residential centers. The program has seen two major evaluations: a thoughtful non-experimental evaluation in the 1970s summarized in PEOPLE (19XX) and an experimental evaluation in the 1990s, summarized in Schochet (2007) et al. The two have remarkably parallel findings; we focus on the experiment.

The first key finding is that removing disadvantaged young men from their local neighborhood dramatically reduces their criminal behavior in the short run. Second, there is a notable effect on educational attainment in the short run, measured in terms of hours, literacy and numeracy and GED and vocational certificate receipt. Third, the Job Corps program generates substantial earnings impacts for 20-24 year old recipients, but not for younger recipients. As a result, because of its high cost, the program does not come close to passing a cost-benefit test (that includes the impacts on crime) for younger participants but does come close for the 20-24 year olds. Despite the lack of an efficiency justification for the program, at least for the 20-24 year olds it actually has a substantial impact on labor market outcomes, which puts it well ahead of many other youth programs, such as JTPA.

WPRS

The Worker Profiling and Reemployment Services (WPRS) system assigns mandatory reemployment services to new Unemployment Insurance (UI) claimants predicted to have long spells of UI receipt or high probabilities of UI benefit exhaustion. A desire to proactively serve likely UI claimants early in their benefit spells, rather than waiting to

serve them until after they have experienced a long spell, motivates the program. The WPRS poses two separate evaluation problems. First, what effect do the mandatory services have on those who receive them and, second, how well does the existing system based on predicted labor market outcomes in the absence of the mandatory services do at allocating them?

We know of two evaluations that address the first question. Dickinson, Decker and Kreutzer (2002) summarize the results of a larger project that includes linear selection-on-observables estimates of the impact of WPRS referral on weeks and amount of UI received as well as earnings and employment for six states. They find substantively important and statistically significant impacts on the UI variables but no systematic effects on labor market outcomes; this suggests that the WPRS system reduces UI usage without imposing a large cost on referred claimants.

More recently, using data from Kentucky and exploiting the particular institutional features of the profiling system in that state, Black, Smith, Berger and Noel [BSBN] (2003) provide experimental evidence of the impact of the reemployment services requirement on claimants who are on the margin for the service requirement given their employment histories and local area characteristics. They find that the program has a substantial effect relative to its (very small) cost, with that effect consisting largely of a deterrent effect, whereby some claimants immediately find employment upon receiving notice of the requirement that they receive services.

BSBN (2003) also address the second question, and find little difference in the impacts by profiling score. Keeping in mind the relative imprecision of their estimates, this suggests that the existing allocation mechanism does not advance economic

efficiency. Pope and Sydnor (2007) argue that the existing mechanism fails on normative grounds as well, though their argument hinges critically on the view that the WPRS treatment represents a burden rather than a benefit.

Employer-Focused Programs

Although it might sound obvious that workforce programs should focus on the labor demand side as well as the labor supply side, until recently there has been a disproportionate emphasis on the latter. In this section we briefly review the literature on three approaches to employer-focused programs: on-the job training (OJT), customized training, and sectoral training.

On-the-Job Training

Subsidized on-the-job training (OJT) at private firms dates back at least to MDTA. This service provides a (typically fifty percent) wage subsidy for a limited period (typically six months) to firms hiring and training certain specified types of workers. Program staff members recruit firms to provide OJT positions (a time-consuming task) and firms always retain the right to reject candidates prior to hiring and to dismiss workers during or after the subsidy period. Though the training provided is supposed to exceed that provided other new workers, anecdotal evidence strongly suggests that OJT recipients often receive the same training as unsubsidized workers (and, in some cases, little or no training at all).

Subsidized OJT has two rationales. The wage subsidy component seeks the purely redistributional goal of getting employers to try out workers who may appear more

risky due to weak labor market histories or other problems. Tying training by the firm to the wage subsidy aims to increase the skills of workers lacking the resources or credit to obtain training either directly from providers or indirectly from firms via lower wages (where the minimum wage may also limit the ability of workers to trade lower wages for training).

Most evaluations suggest positive impacts of OJT on participant employment and earnings. For example, Barnow's (1987) review of the CETA evaluations finds OJT to greater impacts than all other service types. The NJS provides suggestive evidence on this point as well. However, OJT impacts likely embody more displacement than impacts for classroom training and other services that focus exclusively on increasing human capital and not also on redistributing jobs so that partial equilibrium estimates like those noted here do less well at capturing the impacts relevant for a social cost-benefit calculation.

Customized and sectoral training

Customized training is defined as training characterized by: (1) employer input and approval authority for the curriculum; (2) employer authority to establish eligibility criteria for participants and to select participants if they desire; and (3) a commitment by the employer to hire successful program completers. Sectoral training projects occur when customized employment and training services are provided to a group of employers in the same industry or sector of the economy. A number of descriptive studies rhapsodize about these approaches despite the absence of any serious (or even non-serious) impact evaluations. Moreover, the programs lack a clear justification in terms of

economic efficiency; at first blush they seem simply to transfer resources to favored firms or industries under the guise of training. Such schemes may have an efficiency justification if they provide general training to workers who could not finance such training on their own via loans or temporarily lower wages and may have an equity justification if they target disadvantaged groups.

Analytic Issues

This section highlights the four most important analytic issues in the literature. The first concerns heterogeneity in the effects of active labor market policies. This heterogeneity arises both from the fact that programs themselves often provide quite heterogeneous services under headings such as "classroom training". The substantial differences across groups defined by sex and age in average treatment effects noted earlier in our paper strongly suggests that even relatively homogeneous services will have varying effects across individuals as well. In such an environment, evaluation researchers must pay close attention to exactly what average treatment effect their analysis estimates and policy analysts must take care to link the estimates they consider to the policy questions of interest. For example, an experiment with no control group substitution estimates the mean impact of "treatment on the treated". This mean represents the correct impact estimate for a cost-benefit analysis that seeks to address the question of either keeping or scrapping the existing program. It does not provide the correct impact estimate for an analysis of whether the program should receive a larger budget so as to allow it to expand the set of persons served; a simple economic model of program participation suggests that average impacts for individuals on the margin of service receipt will lie below the

mean impact of treatment on the treated.

Second, many studies do not even attempt a cost-benefit analysis, and those that do often provide relatively low quality analyses either due to lack of required inputs and/or failure to follow the best practices outlined in the literature. Without a serious cost benefit analysis, even a relatively strong positive impact estimate has little to say about policy. Without data on all relevant outcomes, as with reliance solely on administrative earnings data for outcomes when programs may also affect, say, criminal behavior and health, means making policy decisions based on incomplete information. Many government programs lack even rudimentary information on either average or marginal program costs, let alone detailed information on the marginal and average costs for particular services and client types. Finally, as noted in HLS (1999), many cost-benefit analyses fail to take full account of the costs of tax funding by omitting consideration of the marginal excess burden of taxation, and proceeding instead as if a dollar of tax funding costs society only a dollar.

Third, most evaluations estimate impacts over relatively short periods from the time of service initiation or random assignment. Recent evidence indicates the dangers this poses to correct inferences about program value. In the negative direction, the early positive impacts found in the National Job Corps Study turned out to largely fade away when longer-term follow up data became available. In the positive direction, classroom training sometimes takes several years to yield its full impact, as in the long-term follow-up of the California GAIN program by HOTZ ET AL and the long term evaluation of German classroom training by LECHNER ET AL. At the same time, the long-term follow-ups of the Supported Work experiment by Couch (1992) and of the JTPA

experiment in GAO STUDY show that sometimes program impacts estimates remain rock solid at the level observed shortly after program participation. With only a handful of studies that provide credible impact estimates more than two or three years out (this paragraph lists nearly all of them), we cannot draw any conclusions about program types or client characteristics associated with long-term impacts.

Finally, only a handful of papers look seriously at general equilibrium effects. Put differently, they ignore the effects that programs may have on the behavior of those who do not participate in them. In addition to indirect effects working through the tax system, these include displacement effects, whereby individuals induced to search harder (or smarter) by a program, or whose skills increase as the result of a program, take jobs that would otherwise have gone to individuals not participating in the program. Programs can also have price effects; for example, a program that produces large numbers of trained auto mechanics or nurses aides should drive down wages in those labor markets. In many cases, failing to take account of general equilibrium effects leads to overly positive conclusions about program performance.

Calmfors (19XX) and JOHNSON provide early conceptual discussions of these issues. The small but growing empirical literature includes Davidson and Woodbury (199X), who find modest but not trivial displacement effects of UI bonuses in a search context. Heckman, Lochner and Taber (200X) find large price effects of a subsidy to university tuition, effects that imply that a partial equilibrium analysis wildly overstates the enrollment effects of the subsidy. Lise, Seitz and Smith (2006) consider the Canadian Self-Sufficiency Project, which provided a generous earnings subsidy to some welfare recipients, and find that taking account of displacement and changes in search effort by

those without the subsidy changes the sign of the cost-benefit calculation for the program.

Conclusions

First, most employment and training programs have either no impact or modest positive impacts. Many if not most do not pass careful social cost-benefit tests, though some that fail may be worth doing on equity grounds. Existing evaluations have important analytic limitations that bias them in favor of programs with short-term impacts and large spillover effects. In general, employment and training programs work best for adult women and least well for youth. The literature provides no good explanation for this demographic pattern.

For reasons of space we have omitted a variety of topics, such as recent studies that examine program design by looking at performance management systems – see e.g. Heckman, Heinrich and Smith (2002) and Barnow and Smith (200X) – at the efficacy of caseworkers – see e.g. Bell and Orr (2002) and ITA EVALUATION (200X) – and at statistical treatment rules as an alternative to caseworkers – see e.g. Eberts, O'Leary and Wandner (2002) and Lechner and Smith (2007). We have also omitted some program categories, such as welfare-to-work programs – see the meta-analyses in XXXX and XXXX – the Trade Adjustment Act, and FILL IN, as well as all evidence from outside the United States – see e.g. Kluve (2007) and WB STUDY. The general lessons from the omitted literature parallel those from what we have covered.

Bibliography

Ashenfelter, Orley. 1978. "Estimating the Effect of Training Programs on Earnings." *Review of Economics and Statistics*. VOL: PAGES.

Barnow, Burt. 1987. "The Impact of CETA Programs on Earnings: A Review of the Literature." *Journal of Human Resources*. VOL: PAGES.

Dickinson, Katherine, Paul Decker and Suzanne Kreutzer. 2002. "Evaluation of WPRS Systems." In Randall Eberts, Christopher O'Leary and Stephen Wandner, eds., *Targeting Employment Services*. Kalamazoo: W.E. Upjohn Institute for Employment Research, 61-90.

Eberts, Randall, Christopher O'Leary and Stephen Wandner. 2002. *Targeting Employment Services*. Kalamazoo: W.E. Upjohn Institute for Employment Research.

Friedlander, Daniel, David Greenberg and Phil Robins. 1997. "Evaluating Government Training Programs for the Economically Disadvantaged." *Journal of Economic Literature* 35(6): 1809-1855.

Heckman, James and Edward Vytlacil. 2007. "Econometric Evaluation of Social Programs." In James Heckman and Edward Leamer, eds., *Handbook of Econometrics, Volume 6*. Amsterdam: North-Holland, forthcoming.

Mueser, Peter, Kenneth Troske and Alexey Gorislavsky. 2007. "Using State Administrative Data to Measure Program Performance." *Review of Economics and Statistics*. 89(4): 761-783.

Perry, Charles, Bernard Anderson, Richard Rowan, and Herbert Northrup, 1975. *The Impact of Government Manpower Programs in General, and on Minorities and Women.* Philadelphia: Industrial Research Unit, the Wharton School, University of Pennsylvania.

Smith, Jeffrey. 2000. "A Critical Survey of Empirical Methods for Evaluating Employment and Training Programs." *Schweizerische Zeitschrift für Volkswirtschaft und Statistik* 136(3): 247-268.

Smith, Jeffrey. 2004. "Evaluating Local Economic Development Policies: Theory and Practice" in Alistair Nolan and Ging Wong (eds.), *Evaluating Local Economic and Employment Development: How to Assess What Works Among Programmes and Policies*. Paris: Organization for Economic Cooperation and Development, 287-332.