

1988

Selection-Bias Correction in an Evaluation of the Vocational Rehabilitation Program

David H. Dean

Robert C. Dolan

University of Richmond, rdolan@richmond.edu

Follow this and additional works at: <https://scholarship.richmond.edu/robins-white-papers>



Part of the [Business Commons](#)

Recommended Citation

Dean, David H. and Robert C. Dolan. 1988. "Selection-Bias Correction in an Evaluation of the Vocational Rehabilitation Program." E.C.R.S.B. 88-8. Robins School of Business White Paper Series. University of Richmond, Richmond, Virginia.

This White Paper is brought to you for free and open access by the Robins School of Business at UR Scholarship Repository. It has been accepted for inclusion in Robins School of Business White Paper Series, 1980-2011 by an authorized administrator of UR Scholarship Repository. For more information, please contact scholarshiprepository@richmond.edu.

SELECTION-BIAS CORRECTION IN AN EVALUATION
OF THE VOCATIONAL REHABILITATION PROGRAM

David H. Dean
Robert C. Dolan
ECRSB 88-8

Selection-Bias Correction in an Evaluation
of the Vocational Rehabilitation Program

David H. Dean

Robert C. Dolan*

Submitted to:

Journal of Human Resources

September 16, 1988

David Dean and Robert Dolan are Assistant and Associate Professors of Economics, Robins School of Business, University of Richmond, Virginia, 23173. This article was funded in part by the University's Faculty Fellowship program.

ABSTRACT

The article applies a "fixed-effect" methodology to evaluate the training impacts of the Vocational Rehabilitation program. Two central issues are identification of an acceptable comparison group and adjusting for sources of selection-bias. The use of program dropouts as a comparison group is examined with a modified Hausman test. The results suggest spending on VR may be more cost-effective in view of the high public cost of serving the disabled in a full dependency mode. These estimates of VR training impacts are compared with the performance of CETA, a public program serving the non-disabled which has been examined extensively using the fixed-effect estimator.

The Vocational Rehabilitation Program (VR) is a federal/state partnership providing services to help persons with disabilities return to work. Since its inception in 1920, the program has been advanced on its economic merit. Today VR is a \$1.5 billion dollar program, a level of public funding suggesting a long history of demonstrated cost-effectiveness. In fact, considerable skepticism shrouds the data and methods which have generated impressive benefit-cost ratios in the past. In the most recent survey, Berkowitz et al. (1988) provide an extensive discussion of the problems facing program evaluation. A major conclusion of that study is the data collected on VR activities do not adequately measure program inputs or client outcomes.¹

The significance of these data limitations is heightened in light of recent developments in the general manpower training literature. Since the Great Society era, federal involvement in employment programs has grown significantly.² Commensurate with this growth have been efforts by economists to measure the earnings effects of government training initiatives. The resulting literature has produced important refinements in the techniques of assessing earnings impacts. These advances include various comparison-group methodologies to control for pre-program differences [Westat, 1982; Dickinson et al., 1986] consideration of "pre-program dip" in program

¹Evaluation is typically conducted using the R-300 data set, a federal reporting system of states' VR activities. While the R-300 reports the total dollar value of services received on a per client basis, there is no clear way to determine the nature, duration, or intensity of treatment. Also, pre-program labor force participation data is scant and no longitudinal information on post-program earnings exists. Finally, no outcome measures other than earnings are collected (e.g. changes in occupation, hours worked, functioning).

²For example, the Manpower Development and Training Act of 1962 (MDTA), the Comprehensive Employment and Training Act of 1973 (CETA), the National Supported Work Demonstration (NSW) of the mid-1970s, and the Job Training Partnership Act of 1982 (JPTA).

evaluation [Ashenfelter, 1975, 1978; Kiefer, 1979; Bassi, 1983] and correction procedures for assorted sources of selection bias in quasi-experimental designs [Heckman, 1979; Bassi, 1984; Ashenfelter & Card, 1985].

This article is the first attempt to apply these assessment methods to the VR Program. For this purpose we have compiled a data set containing longitudinal earnings profiles for trainees. The longitudinal earnings are important because they accommodate econometric tests for selection bias, as well as various specifications to obtain unbiased treatment impacts even in the presence of nonrandom selection. Section I contains discussion of the VR program, its basic data profile and limitations, and the significance of our enhanced earnings profiles. Section II presents the estimation procedure. Section III discusses the choice of a comparison group within the VR setting. Empirical results appear in Section IV.

I. Issues in the Evaluation of VR

The vocational rehabilitation program offers a wide assortment of services to disabled persons. Clients receive varying combinations of diagnosis, restorative medical treatment, education, training, job placement and counseling. This array of service reflects a program which serves persons with physical, mental, or emotional impairments.³ The broad orientation of VR has several implications for evaluation. First, one can expect differential impacts in a training program providing a wide variety of services for a

³The diversity of services and clientele parallels the changing political mandates which have been placed upon the program over its 67 year history. VR was conceived in response to the needs of the physically impaired veterans of World War I. The nature of the VR clientele was expanded in the 1940s with the advent of workers' compensation. VR's mission was again broadened in the Kennedy Administration to serve persons with mental retardation. In the 1970s political emphasis on deinstitutionalization exerted new pressures to treat the chronically mentally ill.

relatively diverse clientele.⁴ Second, the type of disability will often dictate not only the treatment but the treatment impact as well.⁵ VR is also unique in that the service regimen varies greatly in substance and duration across clients.

A state VR agency collects varying degrees of information on each applicant depending on how far the client progresses through the program.⁶ Each client record contains information on demographic traits, type of disability, types of services received, total value of these services, length of time spent in the various treatments, and a limited earnings profile.

Data Limitations

Under current data collection, a client's earnings profile contains a maximum of two earnings points -- at acceptance and closure from the program -- with closure earnings being available only for the fraction of clients completing VR in Status 26. Status 26 denotes a client placed in a job and retaining employment for a period of sixty days. Adopting simplistic assumptions, the net impact of VR services may be calculated as the difference

⁴For an analysis of the differential impacts of the multitude of services provided in CETA see Dickinson et al. (1986).

⁵For example, clients with physical disabilities may have an acute medical condition for which they receive restorative service (prosthesis or surgery). After some education or re-training, they can be placed in a "good" job relatively quickly. In contrast, persons with mental disabilities may have a developmental problem. Their treatment will invariably include work adjustment training for a fixed period and then placement into "sheltered employment" or a low-level competitive position. The emotionally disabled are often clients with a chronic condition. They are usually assigned to a psychiatric-caseload counselor for intensive evaluation, counseling and personal adjustment training, possibly followed by job placement.

⁶Closure statuses for VR clients are defined as follows: Status 08-- Applied but not accepted for services; Status 26 -- Accepted, received services, and rehabilitated; Status 28 -- Accepted, received services, not rehabilitated; Status 30 -- Accepted but no services received (i.e. drop out).

between earnings at acceptance and closure. However, this earnings profile is grossly deficient for evaluation in several respects.

First, the earnings reported at acceptance are unlikely to reflect the true pre-program earnings path of a client due to "pre-program dip". Pre-program dip connotes a decline in client earnings immediately prior to seeking assistance. Although this decline is understandable given that people are more apt to turn to training programs when faced with employment difficulties, it is unlikely that earnings reported at this time capture a trainee's true pre-program earnings potential. The true long-run earnings path may be understated. If so, these earnings do not represent how the client would fare in the absence of treatment, and therefore are a poor baseline for assessing net training effects. Furthermore, VR may be an extreme case of pre-program dip. It is common for clients to report zero earnings in the week prior to application to the program.⁷

A second problem exists with the earnings recorded at closure by Status 26 clients. Although this earnings datum is accurate for the client's sixty days of employment, it is tenuous to impute a post-program earnings path from a single, very short-run employment spell. Indeed, given the rather high recidivism rate in VR, it would be more appropriate to assume decay in the post-program earnings stream.

A third data problem follows from the fact that a significant fraction of clients receiving VR services are closed in Status 28 -- received services but not successfully placed in employment. Obviously no closure earnings are available for this group. However, while it is true that Status 28 clients

⁷Our data indicate that of the cases closed in the Virginia VR program in FY 1982, 78% reported no earnings in the week prior to acceptance.

have no earnings within the limited time perspective of the agency, there is evidence indicating that many clients closed as Status 28 do ultimately get jobs.⁸ The lapse in evaluation is obvious -- Status 28 clients typically receive a substantial level of services and may derive significant benefits from their VR experience, but these benefits will not be captured if employment is not forthcoming within the agency's sixty-day closure vigil.

These specific problems with the client earnings records reflect the more general problem of not having a longitudinal data set. Through the cooperation of the Virginia Department of Rehabilitative Services (VDRS) and the Virginia Employment Commission (VEC), we have constructed a data link through which longitudinal earnings records for VR clients were obtained. The data set contains 9,107 Status 26, 28 and 30 closures from the Virginia VR program for fiscal year (FY) 1982. These clients were matched by social security number with quarterly earnings records of the VEC. The resulting profiles span nine years of quarterly earnings from the fourth quarter of 1976 through the fourth quarter of 1985. This time frame provides sufficiently long pre- and post-program employment histories relative to the 1982 closure data which defines our VR population. Of the 9,107 possible cases, 6,533 had a record of VEC earnings for at least one of the 34 quarters noted above. This reflects a "hit rate" of 72%.⁹

⁸Social Security records for 1977 reveal that the "unsuccessful" Status 28 cases closed in 1975 actually enjoyed average earnings in 1977 of \$3,662. Moreover, these earnings were not dramatically lower than the \$4,041 averaged by Status 26 clients (RSA, 1982).

⁹Matching SSA records encounters coverage gaps. It is impossible to distinguish persons who never worked from those whom either worked in non-VEC covered employment, or who worked "off the books". Also, because the VEC earnings records are state specific, we can not capture the earnings of employed VR clients who move or work out of state.

II. Estimation Procedures

A major issue in recent training program evaluations has been controlling for nonrandom selection bias. Selection bias can surface when the decision to participate in a program is not random with respect to the disturbance term in an earnings function. The problem is that people who undergo a specific treatment regimen may have a different earnings structure from those who do not, even in the absence of training. This bias occurs due to either correlation between the systematic component of the participation equation and the unobserved earnings error (selection on the observables) or correlation between the error terms in the participation and earnings equation (selection on the unobservables components). Much of the recent literature [Bassi, 1983; Ashenfelter and Card, 1985] has focused on the latter relationship.¹⁰ Selection on the unobservables typically stems from either of two sources: 1) self-selection by participants; or 2) "creaming" by program administrators. As Moffitt [1987] has observed, for this selection bias to occur in VR, the client or counselor must have more information on the unobserved error structure than the program analyst. The presence of such a bias will render OLS estimates for the treatment impact inconsistent. More specifically, correlation between the error components of a participation equation and an earnings equation will bias OLS estimates of the treatment impact.

The estimator we apply in this study is concerned with non-random selection based on individual trainee's unobservable characteristics. These are assumed to be constant over time. The conventional form of the model is written:

¹⁰ Bassi's contribution to this evaluation methodology was to derive a more generalized version of the first difference or "fixed effects" model originally developed by Mundlak (1961, 1978) and applied to labor economics by Ashenfelter (1978) and Chamberlain (1982).

$$Y_{it} = X_{it}\alpha + P_i\beta_t + \epsilon_i + \epsilon_t + \epsilon_{it}$$

$$Y_{is} = X_{is}\alpha + \epsilon_i + \epsilon_s + \epsilon_{is},$$

where

Y_{it} = earnings of an individual in the post-treatment year t ¹¹

Y_{is} = earnings for the same individual in the pre-treatment year s ¹²

X_{it} = row vector of human capital characteristics that affect earnings

α = column vector of parameter estimates for these explanatory variables

P_i = binary dummy variable measuring program participation where $P_i = 1$ designates a program participant and $P_i = 0$ if not

β_t = the effect of treatment *for the period t* which can vary for different time periods and thus allows for different payoff profiles¹³

ϵ_i = error component constant over time but specific to the individual, such as IQ or innate ability

ϵ_t = an error component constant across individuals but varying over time, such as macroeconomic phenomena as recessions

ϵ_{it} = white noise error component that varies across time and individuals.

In a random effects OLS specification, the consistency and unbiasedness of the treatment estimator requires that the error component in the equation which is estimated is indeed random. This is seldom true in applications to

¹¹ For the dropouts the t period is for those annual earnings starting with the period immediately proceeding termination from the program.

¹² It is assumed that period s occurs post-disability onset. In the absence of information about the date of disability onset this can only be speculated on.

¹³ The time dependent coefficient on the treatment dummy denotes that the impact of training depends on the date received. With the dependent variable being the level and not the log of earnings the treatment impact coefficient β is not a pure percentage effect. Problems will arise in an inflationary climate where a 5% increase in earnings due to treatment will have a greater impact as earnings go up with inflation. The treatment impact will then be overstated.

training programs. The general fixed effects model advanced by Ashenfelter allows for correlation between the unobserved individual-specific, or fixed component, and the participation dummy. In this framework, a simple differencing of pre and post-treatment earnings regressed on the change in the independent variables between the two periods will yield a consistent estimator of the treatment effect β . The first order difference equation then becomes:

$$Y_{it} - Y_{is} = (\epsilon_t - \epsilon_s) + (X_{it} - X_{is})\alpha + P_i\beta_t + (\epsilon_{it} - \epsilon_{is}).$$

In this model the fixed error component does not vary by assumption and it is simply differenced out.¹⁴

The choice of the appropriate pre- and post-program earnings years is important in the fixed effects difference-in-differences specification. Studies have obtained vastly different treatment effects depending on the choice of base period earnings.¹⁵ In more recent work, Ashenfelter and Card (1985) note that this problem arises because the random component of earnings, ϵ_{it} , may be correlated over time. The presence of pre-program dip indicates that the applicant's transitory component of earnings is unusually low during this time period. Any autocorrelation of the transitory component implies that the participant's earnings also will be low in adjacent periods. Therefore, the difference-in-differences estimates using different base periods of earnings would be expected to yield varying treatment impacts.

¹⁴The Bassi generalization allows for serially correlated errors as well as fixed effects ϵ_i . The problem encountered in adapting this technique to the VR program is that there is no reported information about the date of onset of disability. Subsequently, it is impossible to ascertain whether pre-program earnings are pre- or post-disability onset.

¹⁵Ashenfelter (1978) and Dickinson et al. (1986).

Heckman and Robb (1985) have demonstrated that, despite potential correlation in the random component of earnings, one can still obtain similar estimates of the treatment effect by differencing symmetrically around a decision year. The symmetric difference-in-differences approach accounts for the serial correlation in the transitory component of earnings as long as the appropriate decision year is chosen; thus the main issue is the choice of the appropriate decision year. The difficulty in applying this method to VR is that, unlike most training programs, clients' service regimens vary in duration. The initial model we consider is a first difference fixed effects estimator. We also estimate the model using an earlier pre-program earnings period to account for pre-program dip.

In the fixed effects model, a first-difference of pre- and post-program earnings is regressed on a vector of explanatory variables including a VR participate/dropout (treatment) dummy variable. The specification of the VR earnings model includes change-in-age, change-in-age squared, and a dummy variable for the year of program referral.¹⁶ Human capital theory (Mincer, 1974) predicts that earnings are quadratic in age since age is generally considered as a proxy for experience. However, change-in-age is not as readily interpreted here because the difference in the level of ages is lost in the differencing process. Given the varying lengths of program exposure for VR clients, the change-in-age is not measuring aging effects so much as duration

¹⁶ There are other variables included in previous studies (Dickinson et al., 1986) that may change over the given time interval and thus do not difference out (e.g. marital status, education for younger persons). The cross-sectional R300 data set does not allow the inclusion of such measures.

in the program.¹⁷ As a measure of duration, the change-in-age variable may be picking up two countervailing effects. If the person has low opportunity costs to their time, a longer duration may be expected to have a negative impact on earnings. This would represent the lower value of the opportunity set available to the person. On the other hand, for persons with the same opportunity set, an extra year of treatment may lead to a greater return on the invested time. This may be manifested via a larger earnings level upon program completion and a subsequently positive duration effect.

The year of referral dummies are included to control for business cycle fluctuations (time inhomogeneity). The reference category is referral to the program in 1979. The time dummies for 1980 through 1982 account for economy-wide fluctuations that may have led to the person's applying for services during this period.¹⁸ If the business cycle in Virginia was concurrent with the rest of the country, the economy was in the trough of a recession in 1982. Estimating treatment effects during a recessionary period without controlling for changes in economic conditions will lead to lower estimates than would otherwise be the case. Conversely, the closer the year of referral to the trough of the business cycle, the greater the imputed impact on earnings of remedial manpower training programs. We expect positive signs on the referral year dummies with respect to the reference period of 1979.

¹⁷ By comparing earnings in the pre- and post-training years, everyone's age changes by at least two years. However, VR services are provided to some over a very short period while others stay in the program for several years.

¹⁸ There is no doubt some collinearity between this categorical time dummy and the change in age variables since both measures are reflecting duration in the program. To see this recall that the sample includes all persons closed in fiscal year 1982. All persons referred in 1982 had the shortest time-in-program spell and changed only the minimum of two years in age. For persons referred in 1980, the duration is roughly three years and the change in age will be five years.

III. Issues in Assessing Program Impact in VR

Comparison Group Selection

The optimal methodology for assessing training effects is a pure experimental design in which participants are randomly assigned to treatment and control groups. This approach is infeasible in VR, however, due to the obvious ethical/legal issue of randomly denying the services of a public program to otherwise eligible clients. The next best approach is a "comparison" group methodology.¹⁹

The choice of a comparison group is guided in important respects by institutional aspects of the VR program. The general problem in VR is that enrollment in a particular service regimen is not a random event. A client passes through programmatic and self-screening devices before becoming a VR participant. The client must first apply for the program. This can occur on their own initiative or via referral by another social service agency. Following application, the VR staff makes the decision to accept/reject the person.²⁰ If accepted, a VR counselor suggests a menu of services (IWRP) to be provided. Finally, the client chooses whether or not to participate in the program. Thus an individual who successfully completes the program has made several discrete choices --- to apply for services, to accept the prescribed treatment, and to follow through with the implementation of the IWRP. This process introduces significant selectivity bias since the multistage

¹⁹Discussions of the significance of experimental design versus comparison group methods in training program evaluation are provided by Burtless & Orr (1986) and LaLonde (1986).

²⁰Acceptance to VR requires that: 1) the client have a medically certified physical or mental impairment; 2) the impairment presents a vocational handicap; and 3) the handicap can be remedied through the provision of appropriate services.

participation decision, as well as the subsequent earnings levels of a client, can be systematically influenced by unobserved variables. Minimizing selectivity bias is a major criterion in selecting the comparison group.

For conventional manpower programs it has been common to draw a comparison group from the Current Population Survey (Westat, 1981; Bassi, 1983; Dickinson, et al., 1986). The main problem with the CPS data is there are likely to be significant differences between VR applicants and a random sample of the CPS. For example, even if the observed differences can be controlled for, there can still be unobservable factors such as the initiative which leads a person to apply for VR services. Furthermore, the same unobservable traits that lead a client to participate in VR can also influence their level of success. Therefore, the earnings paths of a CPS sample who have not sought VR services would represent biased estimates of how VR clients may have fared in the absence of training. A comparison group based on the CPS is also flawed due to insufficient information about the presence of a disability for persons in the survey. This is critical since disability is a major factor determining an individual's pre-program earnings profile. Different types of disabilities can influence pre-program earnings, the service regimen provided by the VR agency, as well as the outcome of these services. For these reasons the CPS is inappropriate for a comparison group for evaluation of VR clients.

The alternative to the CPS is an "internal" comparison group -- a cohort having had some degree of exposure to the program. The institutional nature of VR offers three different classes of internal comparisons groups: 1) persons who applied for but were not accepted for VR services (Status 08); 2) persons not successfully rehabilitated after receiving services (Status 28); and persons who are accepted and agree to participate, but leave the program

prior to receiving any treatment (Status 30).

Several studies of VR [RSA, 1982; Abt & Associates, 1974] have used the Status 28 cohort (not rehabilitated) as a comparison group. Since this group passes through the multiple screens to become a program participant, it would seem that selection bias would not be an issue. However, the Status 28 cohort is unacceptable because this group suffers from a variation of "contamination bias" -- they have been exposed to the services that normally would be reserved for the treatment group in a true experimental setting. Accordingly, one lacks the dichotomization necessary to isolate the effects of the intervention undertaken. Indeed, one should not only avoid the Status 28's as a comparison group, but rather include them as part of the treatment group. Omitting this cohort from the treatment group imparts an clear upward bias to the service impacts since only "successfully" rehabilitated clients will remain.

The second possible internal comparison group contains individuals declared ineligible for VR services (Status 08).²¹ We reject this group because they differ in both observable and unobservable characteristics vis-a-vis the treatment group. The problem is that acceptance involves elements of self-selection and programmatic screening. Moreover, the programmatic screens can imply informal consideration by administrators such as "creaming" or even "scraping".²² The latter may be especially true in VR since the program was mandated to serve a more severely disabled population in 1973. These measured and unmeasured differences preclude them as a proxy for what the earnings of VR participants would have been in the absence of treatment.

²¹Status 08 are persons judged by the VR staff to not have a disability or vocational handicap, or persons with disabilities deemed irredeemable by VR. The deficiencies of the Status 08 are discussed by Englander (1985).

²²These issues are addressed by Bassi (1983) and Card & Sullivan (1986).

We believe the criterion of minimizing pre-enrollment differences recommends Status 30 clients as the comparison group. Recall, the Status 30 client is a dropout. The use of dropouts may invite negative connotations initially; in fact, this cohort has conceptual appeal.²³ First, Status 30 clients had the motivation to apply and they met the eligibility criteria. This obviates the need to model the application and acceptance decision. Second, it is likely that both successfully rehabilitated clients and dropouts experience similar depreciations in human capital which lead them to apply to VR. Finally, because the Status 30 clients are closed without receiving any prescribed treatment, the only service received by this cohort is a diagnostic evaluation. Diagnosis alone is unlikely to have significant effect on the client's human capital.

Still, the fact that the Status 30 clients self-select out of the program at an early stage suggests unobservable attributes within this group which can introduce bias. Is this group generally less motivated? Or is it more ambitious, viewing the employment prospects on their own as greater than with VR assistance?²⁴ Or has the employment outlook for this group systematically improved relative to what it had been? The concern is that some elements of unobservable difference between the treatment and comparison groups which are correlated with future earnings probably persist. However, on balance, the measured and unmeasured difference between the Status 30 cohort versus any other viable comparison group would surely be greater.

²³Cooley et al. also conclude that "no-shows" are a superior comparison group in many training settings because "they have many characteristics in common with those who were trained" (1979; pp. 123-124).

²⁴ Burtless (1985) discovered that some graduates of trainee programs were adversely stigmatized in the eyes of employers due to participation.

Testing the Quality of the Internal Comparison Group

Bassi's model tests specifically for the presence and significance of "creaming" on the part of program administrators. Creaming implies that program administrators, who are evaluated on the basis of successful closures, will accept persons for services who are temporarily below their permanent earnings path. In doing so they are selecting clients likely to fare the best since they have higher permanent income levels.²⁵ The concept of creaming is given a unique twist in VR program due to the 1973 mandate that the program accept persons with severe disabilities. On the one hand, this mandate connotes a "scraping" rather than creaming bias for some clients. At the same time, however, counselors may feel even greater pressure to cream when possible in order to compensate for the more severely disabled clients in their caseload who are more difficult to bring to a successful closure.

Happily, the significance of creaming or scraping is minimized in a comparison group comprised of program dropouts by virtue of the fact they were accepted for services under the same criteria as the treatment group. Furthermore, creaming can be controlled for if the base period chosen is one several years prior to application for services. We can also test and control for correlation between the participation decision and individual time-specific earnings error component in the years prior to being accepted.

A more significant concern when using dropouts as a comparison group is to determine whether there are differences in the earnings structures of the groups with respect to the transitory nature of pre-program dip. The

²⁵Selection bias occurs if the earnings component of one of the groups experiences a transitory rather than permanent decline in the period immediately prior to application to the program. This situation underscores the importance of adequately adjusting for "pre-program dip".

two groups may have similar earnings structures for several years prior to program application, and then differ only in the period just prior to program participation. Such a circumstance reduces the reliability of the measured treatment impacts since a portion of the observed earnings gain is merely a return to the permanent earnings path from which the person had temporarily fallen. If the decline in earnings is largely permanent for the treatment group, then using a base period several years prior to program participation will underestimate the treatment impact. Since we are unable to untangle permanent from transitory declines in the earnings profiles, by using various base periods, we obtain upper and lower bounds of the treatment effect.²⁶

To examine the validity of the Status 30's as a control for selection bias in a fixed effects framework, we use a modified Hausman Test of unobservables.²⁷ The test focuses on a VR participation dummy in a first-differencing equation using *pre-program* earnings. The dummy measures the mean value of the error term -- in this instance, the error term differenced over the two periods. A significant coefficient indicates selection bias due to correlation between the error term and program participation. Since we have only three years of pre-program earnings for the majority of the sample, there are two dependent variables to consider -- change in earnings over periods (s-1) and (s-2) and (s-2) and (s-3). Significant coefficients

²⁶ Unfortunately, we do not have data on the date of onset of the disabling condition. If we did, we could examine the earnings streams both pre and post-disability onset to determine whether the dip is permanent or transitory.

²⁷ We also tested the assumption that the observable component of the earnings functions for the treatment (Status 26 & 28) and comparison groups (Status 30) would be the same in the absence of VR. Using pre-program earnings for the various years as the dependent variable, a Chow test reveals if the coefficients for a random effects (OLS) estimator differ between the treatment and dropout groups. Of the six disability cohorts, the F-statistic for equality of coefficients was significant only for physically disabled men.

on the participation dummies suggest unobservable differences between the treatment and comparison groups and thus render the fixed effects estimation inconsistent.²⁸

The results of the Hausman tests are reported in Table 1. Estimation of the earnings equation was performed by disability (physical, mental and emotional) and gender stratifications.²⁹ The coefficients on the treatment variable were significant (and positive) for physically disabled men in both intervals. This implies pre-program unobservable differences for these clients, making their pre-program earnings significantly higher than program dropouts. The Status 30 cohort is thus a poor comparison group for men with physical disabilities. The treatment dummy was also significant for physically disabled women in the first pre-program interval, but insignificant in the interval for two to three years prior to program referral. For each of the remaining four cohorts there were no significant unobservable differences between cohorts for either of the two time intervals. Therefore, except for physically disabled men, a fixed effects estimator should lead to unbiased and consistent estimates if we use earnings in either the one or two years prior to program referral as the base year for measuring treatment impacts.

Longitudinal Data Alignment Issues

Our VR data set is comprised of closed cases from Virginia's program during FY 1982. When comparing service-receiving clients with program "drop-outs", one must consider the likelihood that these cohorts came to the

²⁸ When Bassi found such evidence she concluded that for white males the selection-bias was so strong that even a fixed effects specification would lead to biased treatment effects. She excluded this cohort.

²⁹ Noted that the change in age variable in a first differencing reduces to a vector of 1's and therefore is dropped from the estimation procedure.

TABLE 1

Modified Hausman Test Results for a Fixed Effects Specification
(Sorted by Disability Type for Two Pre-Program Earnings Intervals)

Variables	Physically Disabled				Mentally Disabled				Emotionally Disabled			
	Dependent Variable:				Dependent Variable:				Dependent Variable:			
	Prior Year 1 Earnings Minus Prior Year 2 Earnings	Prior Year 2 Earnings Minus Prior Year 3 Earnings	Prior Year 1 Earnings Minus Prior Year 2 Earnings	Prior Year 2 Earnings Minus Prior Year 3 Earnings	Prior Year 1 Earnings Minus Prior Year 2 Earnings	Prior Year 2 Earnings Minus Prior Year 3 Earnings	Prior Year 1 Earnings Minus Prior Year 2 Earnings	Prior Year 2 Earnings Minus Prior Year 3 Earnings	Prior Year 1 Earnings Minus Prior Year 2 Earnings	Prior Year 2 Earnings Minus Prior Year 3 Earnings	Prior Year 1 Earnings Minus Prior Year 2 Earnings	Prior Year 2 Earnings Minus Prior Year 3 Earnings
Intercept	653.53 (1.54)	703.98 (2.28)*	1156.65 (2.36)*	864.67 (2.89)**	665.35 (2.68)**	659.61 (2.23)*	701.85 (2.93)**	242.23 (1.06)	1060.97 (2.24)*	238.95 (0.54)	1174.61 (2.24)*	1534.33 (3.48)**
Treatment	729.96 (2.74)**	455.94 (2.19)*	929.42 (3.00)**	185.42 (0.91)	250.71 (1.48)	-146.68 (0.72)	-372.41 (2.25)*	-104.76 (0.65)	277.13 (1.09)	381.70 (1.35)	73.12 (0.26)	-201.63 (0.71)
Referred in 1980	-1533.23 (4.88)**	-803.34 (3.32)**	-955.96 (2.62)**	-231.22 (0.98)	-220.18 (1.13)	-242.24 (1.28)	-249.44 (1.31)	126.57 (0.86)	-742.80 (2.31)*	-367.68 (1.35)	-614.18 (1.70)	-615.16 (2.25)*
Referred in 1981	-1050.88 (3.52)**	-330.15 (1.46)	-1904.8 (5.48)**	-461.59 (2.10)*	-321.71 (1.70)	-414.17 (2.14)*	-644.95 (3.48)**	-60.79 (0.40)	-1088.69 (3.53)**	-732.03 (2.68)**	-1216.07 (3.50)**	-1135.72 (4.13)**
Referred in 1982	-74.86 (0.18)	-21.18 (0.08)	-2153.93 (4.38)**	-755.67 (2.91)**	-834.81 (3.11)**	-642.60 (1.61)	-503.77 (1.92)	-545.78 (1.74)	-512.09 (1.08)	-628.17 (1.28)	-1377.07 (2.57)*	-1554.16 (3.14)**
Change in Age Sq.	19.52 (4.96)**	13.20 (5.10)**	1.19 (0.78)	-0.43 (0.51)	18.10 (5.67)**	4.33 (1.14)	1.56 (1.50)	0.78 (0.79)	19.11 (3.36)**	8.67 (1.82)	-0.06 (0.03)	0.92 (0.58)
R-Square	0.03	0.03	0.02	0.01	0.05	0.01	0.02	0.01	0.04	0.03	0.02	0.03
F Ratio	13.69	9.90	9.96	2.55	11.27	1.54	3.46	1.30	5.96	3.29	3.38	4.16
Number of Obs.	1985	1516	1985	1516	1094	587	1094	587	747	604	747	604

† Ratios in Parentheses

* Denotes significance at the .05 level
**Denotes significance at the .01 level

program at slightly different times. It is evident from Table 2 that the treatment groups entered the VR program earlier -- generally the latter part of 1980 -- as compared to the comparison groups' entry in early to mid 1981. Only the treatment and comparison cohorts for the physically disabled entered the program at roughly the same time. The fact that the treatment cohorts spend time in the program means they are receiving services while their comparison group has yet to apply. One must adjust for this timing lapse in order to control adequately for pre-program dip.³⁰ The appropriate pre-program earnings points must be the same period for both treatment and comparison groups, and must be far enough prior to both groups' application to the program so that neither group is likely to have earnings data reflective of pre-program dip.³¹ The problem is the dip occurs uniformly earlier for the treatment group than the comparison group. For example, while the comparison group probably "dips" in 1980 (depending on actual onset of disability), the treatment group -- many of whom are already in the program -- will have likely experienced their pre-program dip in late 1979 or early 1980. For this reason, examining the longitudinal earnings data on a calendar year basis is inappropriate.

Since termination from the program can occur at any point over a 12 month period, there is also an alignment issue in determining the appropriate post-program period. If a calendar quarter approach is taken, it is conceivable there will be comparisons between groups which entered the labor force three quarters apart. This seriously biases the estimated training impact if there is time inhomogeneity (e.g. inflation).

³⁰ This "alignment" problem is discussed extensively by Bassi (1984) and Dickinson et al. (1986).

³¹ This "sampling frame" problem is discussed by Dickinson et al., (1986).

TABLE 2

Comparing Program Duration of Clients Terminated from the VDRS in 1982

Male Clients									
Physical Disability				Mental Disability			Emotional Disability		
Variable	Treatment (n=1686)	Comparison (n=299)	T Statistic	Treatment (n=903)	Comparison (n=191)	T Statistic	Treatment (n=614)	Comparison (n=133)	T Statistic
Average Duration in Months	5.01	4.67	2.22 *	5.37	3.99	6.87 **	5.31	4.01	5.96 **
Quarter Referred to VR Program	1st, 1981	1st, 1981	1.43	4th, 1980	2nd, 1981	6.60 **	4th, 1980	2nd, 1981	5.75 **
Female Clients									
Physical Disability				Mental Disability			Emotional Disability		
Variable	Treatment (n=1362)	Comparison (n=154)	T Statistic	Treatment (n=509)	Comparison (n=78)	T Statistic	Treatment (n=522)	Comparison (n=82)	T Statistic
Average Duration in Months	4.24	4.12	0.55	6.15	3.99	7.84 **	5.70	3.80	6.90 **
Quarter Referred to VR Program	1st, 1981	1st, 1981	0.04	3rd, 1980	2nd, 1981	7.57 **	4th, 1980	2nd, 1981	5.90 **

* Denotes Significance at the .05 Level
** Denotes Significance at the .01 Level

The importance of proper earnings alignment for a program of varying duration is further demonstrated in Table 2. The table reports the quarter in which a cohort, on average, was referred to VR and the average duration of treatment. The table is stratified by disability, gender, and treatment status. Note that the mentally retarded and emotionally disabled treatment cohorts average roughly two more months in the program and are referred for VR two quarters before their Status 30 (dropout) counterparts. In contrast, the physically disabled treatment and comparison cohorts are referred in the first quarter of 1981 and spend about the same amount of time in program.

These alignment difficulties are peculiar to VR because of differing service durations. To address the issue, the earnings data were reconstructed into intervals corresponding to the appropriate pre- and post-program periods. Given the varying times of entry and closure, it is more accurate to use the earnings for the periods that start in the quarters immediately prior to acceptance and after treatment. Thus the sum of earnings reported in the four quarters immediately prior to program acceptance represent Year 1 pre-program earnings and the four quarters following closure comprise Year 1 post-program earnings. Annual earnings were similarly determined for the corresponding pre- and post-program periods for Year 2 and Year 3.³²

Earnings summaries using annual earnings prior to referral and post closure for the male and female cohorts are reported in Table 3. T-tests for significant difference between the stratified comparison and treatment groups

³² There are missing earnings data for three quarters of 1979 so for some clients the four quarters grouped together for the pre-program year's earnings may not be consecutive quarters. This gap is less serious if it covers the period with in which many of the clients became disabled. We have no way of knowing, however, because VR agencies do not request information on the date of onset of the client's disabling condition.

TABLE 3

Comparing Pre- and Post-Program Earnings of Clients Terminated from the VDRS in 1982

Variable	Male Clients			Female Clients		
	Physical Disability		T Statistic	Mental Disability		T Statistic
	Treatment (n=1686)	Comparison (n=299)		Treatment (n=903)	Comparison (n=191)	
3 Year Prior Earnings	\$3,066	\$3,883	2.36 *	\$1,075	\$1,051	0.12
2 Year Prior Earnings	\$4,019	\$3,823	0.56	\$1,242	\$1,468	0.93
1 Year Prior Earnings	\$3,107	\$2,103	3.76 **	\$961	\$926	0.21
1 Year Post Earnings	\$4,110	\$2,685	3.69 **	\$2,109	\$1,870	0.81
2 Year Post Earnings	\$4,591	\$2,939	4.47 **	\$2,535	\$2,923	0.78
3 Year Post Earnings	\$4,862	\$3,082	4.72 **	\$2,887	\$3,114	0.59
Variable	Physical Disability		T Statistic	Mental Disability		T Statistic
	Treatment (n=1362)	Comparison (n=154)		Treatment (n=509)	Comparison (n=78)	
3 Year Prior Earnings	\$1,758	\$1,523	1.12	\$406	\$544	0.82
2 Year Prior Earnings	\$2,316	\$1,908	1.67	\$679	\$813	0.64
1 Year Prior Earnings	\$2,176	\$1,384	3.57 **	\$709	\$880	0.70
1 Year Post Earnings	\$3,287	\$1,494	6.76 **	\$1,755	\$765	3.96 **
2 Year Post Earnings	\$3,416	\$1,672	5.54 **	\$2,099	\$817	4.37 **
3 Year Post Earnings	\$3,557	\$1,902	4.80 **	\$2,167	\$1,164	3.13 **
Variable	Physical Disability		T Statistic	Emotional Disability		T Statistic
	Treatment (n=1686)	Comparison (n=299)		Treatment (n=903)	Comparison (n=133)	
3 Year Prior Earnings	\$3,066	\$3,883	2.36 *	\$1,753	\$2,105	1.01
2 Year Prior Earnings	\$4,019	\$3,823	0.56	\$2,146	\$2,238	0.26
1 Year Prior Earnings	\$3,107	\$2,103	3.76 **	\$1,619	\$1,329	1.12
1 Year Post Earnings	\$4,110	\$2,685	3.69 **	\$2,371	\$1,733	1.60
2 Year Post Earnings	\$4,591	\$2,939	4.47 **	\$2,485	\$1,795	1.59
3 Year Post Earnings	\$4,862	\$3,082	4.72 **	\$2,697	\$1,892	1.81

* Denotes Significance at the .05 Level

** Denotes Significance at the .01 Level

are also presented. There are several striking comparisons to be drawn. First, in almost all cases we observe a significant pre-program dip in earnings for the first year prior to program application compared to the second year prior to application for services. The change in earnings from the third to second year prior to program referral was much smaller and in several cases earnings increased during this interval.³³ The post-program earnings for the treatment group were always higher than the comparison group, with the exception of the later years' earnings for mentally disabled men. These differences were significant for the male and female physically disabled cohorts as well as for the female mentally retarded cohort. In terms of inter-disability comparisons, the physically impaired earned more than the emotionally disabled who in turn had higher earnings levels than the mentally retarded. The male-female earnings differentials depended on the type of disability.

IV. Empirical Results for Fixed Effects Earnings Impacts

It is clear from the Hausman tests that a fixed effects model is the preferred estimator. Using this framework, we estimate VR treatment impacts based on earnings two years prior to application as the base year and one year after closure as the outcome year. The base year is chosen in deference to the pre-program dip for the dropout comparison group. Recall, the Hausman test indicated the biases encountered with using the first pre-program year earnings as the base period.³⁴

³³ This may reflect the onset of disability rather than the voluntary withdrawal from the labor market which pre-program dip often connotes. If so, the more recent earnings represent the client's new permanent earnings path.

³⁴ Also recognize that the period around which earnings are compared is the interval, not the year, of program participation. Without a fixed starting date for the treatment and comparison group, and considering the varying duration of VR service regimens, the choice of a uniform base year is inappropriate.

Results for six gender/disability stratifications appear in Table 4. Observe that the treatment coefficients are positive for all six cohorts, although statistically insignificant in two cases (mentally disabled men and emotionally disabled women). The significant treatment coefficients range from \$727 for emotionally disabled men to \$1578 for physically disabled women.

The year-of-referral dummies are generally insignificant except for mentally retarded women for 1981 and 1982. Indeed, for this cohort the year-of-referral coefficient is greater than the treatment dummy. However, observe how this coefficient becomes increasingly negative (albeit insignificant) the shorter a person is in the program. This result suggests this variable may be picking up more of a duration effect than the impact of economy-wide fluctuations.³⁵

The change-in-age variable is positive and significant for three of the six cohorts. The coefficients are also quite large, larger than the treatment effect in two cases. This result may be due to the relatively young age of the cohorts. Given the lack of work experience for many clients, a large impact of change-in-age is not surprising. As expected, the quadratic age specification is negative and significant for four cohorts. But this variable is probably collinear with the referral period dummies in that, given a fixed point of closure, both variables are proxies for program duration. This makes the interpretation of change-in-age in a VR setting less clear than in a program of fixed duration.

³⁵ Recall that 1982 was the trough of the recession and it may have been that these persons coming to the program at this time were experiencing cyclical unemployment. The program then served as a placement device for persons with relatively minor impairments, at least those which did not require a lengthy stay in the program as evidenced by the short interval from referral to closure.

TABLE 4
Fixed Effects Earnings Estimates Sorted by Impairment and Gender
(VDRS Clients Closed During Fiscal Year 1982)

Dependent Variable: Change in earnings between two years prior to acceptance and one year after closure

Variables	Physically Disabled		Mentally Disabled		Emotionally Disabled	
	Male	Female	Male	Female	Male	Female
Intercept	-5707.37 (2.36)*	-3096.57 (1.65)	-4241.35 (2.04)*	3718.58 (1.76)	-2943.42 (1.08)	-1216.55 (0.42)
Treatment ^a	1212.49 (2.95)**	1578.57 (4.44)**	400.32 (1.17)	890.05 (2.18)*	727.67 (1.67)*	734.32 (1.41)
Referred in 1980	-517.52 (0.92)	82.30 (0.17)	623.53 (1.33)	-444.54 (1.03)	8.84 (0.01)	-36.05 (0.06)
Referred in 1981	54.27 (0.07)	203.59 (0.31)	835.86 (1.21)	-1482.43 (2.23)*	-176.62 (0.19)	-97.20 (0.10)
Referred in 1982	1542.87 (1.45)	252.78 (0.31)	1225.74 (1.34)	-2523.59 (2.33)*	740.25 (0.59)	-258.56 (0.19)
Change in Age	1995.04 (4.22)**	1072.63 (2.95)**	1111.33 (2.80)**	-582.23 (1.40)	1026.46 (1.90)	385.41 (0.65)
Change in Age Sq.	-24.56 (9.23)**	-13.60 (6.54)**	-6.22 (2.17)*	-0.91 (0.30)	-14.18 (3.33)**	-5.07 (1.40)
R-Square	0.05	0.04	0.01	0.03	0.02	0.01
F Ratio	17.72	10.01	2.73	2.79	2.90	0.81
Number of Obs.	1985	1516	1094	587	747	604

T Ratios in Parentheses
a Denotes one-tail test reflecting the predicted impact of treatment effects.
* Denotes significance at the .05 level.
**Denotes significance at the .01 level.

The F-test for women with emotional disabilities indicates an overall lack of explanatory power of the model. However, this result is perhaps unsurprising given the nature of the disability. Of the three impairment classifications, emotional conditions are most likely to deteriorate over time. If this is occurring during treatment, the impact of VR services could be offset by a decline in functional capacity. If functioning tends to change significantly, it would be desirable to include a health variable as a control. Recall that in our present fixed effects specification, factors such as health are assumed to be invariant or change very slowly with the passage of time.

A second caveat to these findings stems from a lack of information about the client's earnings path before and after the onset of disability. Without this information we cannot be certain we have identified the appropriate baseline earnings. The estimates reported above use earnings for the period two years prior to treatment. In doing so, we are attempting to identify baseline earnings reflective of the client's earnings path prior to the onset of disability. This focus could be inappropriate. In the case of a traumatic accident, it is conceivable that, in the absence of treatment, the person's post-disability earnings potential may indeed be zero. If so, post-disability earnings are the relevant baseline. For this reason, it is reasonable to think of the two year pre-program earnings path as rendering a lower bound estimate of treatment impacts. Re-estimation of the model using earnings at referral would yield an upper bound estimate.

Despite these qualifications, the magnitude of the treatment impacts we find for VR are comparable to those of the CETA program in 1977. Several studies (WESTAT, 1982; Bassi, 1983) found the training impacts for women were generally positive and significant depending on the type of program and

minority status. The earnings gains ranged from \$550-1200 for on-job-training (OJT), \$500-550 for classroom training, and \$650-950 for public sector employment. The treatment impacts for work experience, which is similar to VR's work adjustment training provided to mentally retarded persons, were insignificant. Classroom training and OJT are similar to the services provided to the physically disabled who, in addition, receive some restorative services. For men, the CETA program impacts are often found to be negative. This is explained by the loss of work experience while enrolled in the program. Thus the findings of positive and significant coefficients for the male physical disability group receiving VR services presents a marked contrast with the more recent manpower training programs serving the economically disadvantaged.

V. Conclusions

While the VR program consumes \$1.5 billion annually, it represents only five percent of the public transfer payments going to persons with disabilities. This fact prompts some obvious questions. Is public policy giving appropriate emphasis to rehabilitation and training in the mix of programs assisting persons with work disabilities? Should VR be expanded, perhaps at the expense of income transfer or other direct services, or instead, does efficiency suggest the program be scaled down. The answers to these questions depend on researchers' abilities to measure the earnings impacts of VR. As one of the oldest federal remedial manpower training initiatives, VR has not been without numerous evaluations; but the reliability of the estimates is vastly diminished by the quality of both the data and methods employed. The most glaring shortcomings are the lack of a randomized control group and the shallow earnings profiles typically available through the state agencies.

With the acquisition of longitudinal earnings for clients of Virginia's VR program, we have used evaluative techniques which have not been applied to VR. Lacking a pure experimental design, we addressed selection bias using a defensible, though still imperfect, internal comparison group. Using a fixed-effects framework, we found the treatment impacts of VR to be significant for four of six cohorts. Significant earnings gains ranged from \$727 to just under \$1,600 for the period -- estimates which appear plausible compared to those reported for CETA.

However, our confidence in these findings is compromised by two data issues unique to evaluating a program serving the disabled. First, our fixed effects specification should be expanded to include a variable to control for the possibility that the client's health status changes during treatment. If health changes significantly, this imposes bias since these treatment impacts are estimated under the core assumption that unobservables are fixed or change only slowly with time. Second, we need more precise information about the date of disability onset. This would provide a clearer basis for identifying a base earnings year reflective of the client's true earnings path.

Bibliography

- Ashenfelter, Orley. 1978. "Estimating the Effect of Training Programs on Earnings." Review of Economics and Statistics 60:47-57.
- Ashenfelter, Orley and David Card. 1985. "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs." Review of Economics and Statistics. 648-660.
- Bassi, Laurie J. 1984. "Estimating the Effect on Training Programs with Non-Random Selection." The Review of Economics and Statistics 66.1:36-43.
- Bassi, Laurie J. 1983. "The Effect of CETA on the Postprogram Earnings of Participants." The Journal of Human Resources 18.4:539-556.

- Berkowitz, Monroe et al. 1988. Analysis of Costs and Benefits in Rehabilitation. Temple University Press. Philadelphia
- Dickinson, Katherine, Terry Johnson, and Richard West. 1986. "An Analysis of the Impact of CETA programs on Participants' Earnings." The Journal of Human Resources 21: 64-91.
- Fraker, Thomas and Rebecca Maynard. 1986. "The Use of Experimental Versus Comparison Group Designs in Evaluations of Employment-Related Programs." mimeo. Mathematica Policy Research, Inc.
- Heckman, James. 1979. "Sample Selection Bias as a Specification Error." Econometrica 47:153-161.
- Heckman, James and Richard Robb. 1985. "Alternative Methods For Evaluating the Impact of Interventions." In Longitudinal Analysis of Labor Market Data, ed. J. Heckman & B. Singer. Cambridge: Cambridge University Press.
- Kiefer, N. 1978. "Federally Subsidized Occupational Training and the Employment and Earnings of Male Trainees." Journal of Econometrics 8:111-125.
- Kiefer, Nicholas. 1979b. "Population Heterogeneity and Inference from Panel Data on the Effects of Vocational Training." Journal of Political Economy 87.5 (part 2):S213-S226.
- LaLonde, Robert J. 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." American Economic Review 76:4, 604-620
- Lewis, H. Gregg. 1974. "Comments on Selectivity Biases in Wage Comparisons." Journal of Political Economy 82:1145-1155.
- Maddala, G.S. 1983. Limited Dependent and Qualitative Variables in Econometrics. Cambridge: Cambridge University Press.
- Mincer, Jacob. 1974. Schooling, Experience, and Earnings. New York: National Bureau of Economic Research.
- Nowak, Laura. 1983. "A Cost-Effectiveness Evaluation of the Federal/State Vocational Rehabilitation Program--Using a Comparison Group." The American Economist 27:23-29.
- Thornton, Craig, Peter Kemper and David Long. 1980. "A Benefit-Cost Analysis of the Supported Work Experiment." In Public Expenditure and Policy Analysis, ed. by R. Haveman and S. Margolis, New York: Houghton Mifflin.
- Worrall, John D., "A Benefit-Cost Analysis of the Vocational Rehabilitation Program." Journal of Human Resources 13.2: 285-298.