

APPENDIX

D

Estimates of Effects of Employment and Training  
Programs Derived from National Longitudinal  
Surveys and Continuous Longitudinal Manpower Survey

*Valerie Nelson and Charles F. Turner*

In addition to the program-specific evaluations of YEDPA effectiveness reviewed in Chapters 4 through 8, several evaluations have used large representative samples of the American population to derive estimates of the overall impact of all federally funded employment and training programs. The most prominently used data bases in those studies are the Continuous Longitudinal Manpower Survey (CLMS, administered by Westat with data collection by the U.S. Bureau of the Census), and a special youth sample of the National Longitudinal Survey (NLS, administered by the Center for Human Resource Research of Ohio State University with data collection by the National Opinion Research Center--NORC). Both of these surveys involve relatively large samples (over 60,000 in the CLMS and over 12,000 in the NLS) drawn in a manner designed to permit generalizations to the universe of American youths (NLS) or participants in CETA programs (CLMS). (It should be noted that only a fraction of the youths in the NLS sample participated in federally funded employment and training programs, and similarly, only a fraction of the program participants sampled in the CLMS survey were youths.)

While the major charge of our committee was to focus on the Youth Employment and Demonstration Projects Act (YEDPA) knowledge development activities, it seemed prudent to review the findings from studies using these other data bases. Since these studies use data gathered in a different manner and have a somewhat different (and wider) focus, they provide an important supplementary perspective on the substance and problems of the individual YEDPA evaluations we reviewed. Moreover, because these studies use data derived from samples with high sample-coverage rates and low sample attrition, they can provide a more adequate evidentiary basis (at least in respect to sampling methods) than many of the other studies we reviewed.

There are, nonetheless, important limitations to these data bases, as well. First, they are not targeted on specific programs--and so the relevant estimates of aggregate program effects may lump together both

---

Valerie Nelson, an economist, was a consultant to the committee. Charles F. Turner was senior research associate with the committee.

effective and ineffective programs. Second, the data bases (particularly CLMS) limit the extent to which one can take into account the effects of local labor market conditions. Third, these data are not derived from experiments in which subjects were randomly assigned to take part in a program, and hence the resultant estimates of program effectiveness require strong assumptions about the adequacy of the model specification and the matching procedures used to construct synthetic "control" groups. Finally, we should point out that the CLMS reports were provided to the committee in "draft" form late in the course of our work, and thus our evaluation of them has not been as intensive as that of the individual YEDPA reports.

In the following pages we briefly review the characteristics of the NLS and CLMS data bases. We then describe the findings of empirical studies which have used these data bases to estimate program effectiveness. We conclude by discussing the most serious problem with all of the studies (and many of the previously reviewed YEDPA studies): potential biases in the selection of a "control" group of nonprogram participants.

#### CHARACTERISTICS OF CLMS AND NLS DATA BASES

Both the CLMS and the NLS are full probability samples whose sampling designs appear to have been well executed. That is to say, sample coverage appears high and the available documentation shows that considerable attention was given to important methodological details, such as adequacy of sampling frame, careful screening of respondents to ensure that they fell within the universe being sampled from, extensive follow-up to ensure a high response rate, and so on.

The CLMS was designed to sample all persons receiving training or employment under the Comprehensive Employment and Training Act (CETA), and of course, a portion of that sample would have been young people. (The CLMS sample of program participants is complemented by data for nonparticipants derived from the Current Population Survey (CPS).) The NLS youth sample, on the other hand, is a longitudinal survey of American youths begun in 1979. Because it sampled the entire population of young persons (and indeed oversampled groups who were likely to participate in employment and training programs), it does include a sample of young persons who happened to have participated in YEDPA or CETA programs.

#### Sample Execution

Rough lower bounds on the number of program participants aged 18-21 in each sample are 21,000 for the CLMS and 1,800 for the NLS. According to information provided by the NLS Data Center at Ohio State University, the NORC sampling report for NLS surveys (Frankel, McWilliams, and Spencer, 1983), and the published CLMS estimates (for initial waves of data collection), it appears that sample attrition rates were on the order of .05 to .10 per wave in both surveys. Because the NLS retained

in the sample persons who did not respond in the previous wave of data collection, it is said that each of the 1980-1983 waves interviewed almost .95 of the wave-1 sample. The CLMS, in contrast, lost about .10 of its sample per wave and did not retain nonrespondents from previous waves. Thus by wave 4, it obtained interviews with only .729 of the wave-1 sample.

#### Measurements

To the extent one focuses on labor force outcomes and income as dependent variables and uses standard human capital-type variables (e.g., education, training, age), either data base might be of value, although the lengthy NLS questionnaire contains a much wider range of social and economic measures than does the CLMS. (One must wonder, however, about the extent to which fatigue may contaminate the replies of the respondents to the NLS.)

#### Error Structure

The reliability and validity of these data do not appear (from the available documentation) to be buttressed by explicit estimates of the error and bias introduced by respondents, interviewers, coders, processors, and so on. Since the Census Bureau has a routine re-interview procedure, some test-retest measurements are likely available for the CLMS data set. Information on other aspects of the error and bias that affect these data do not appear to be available. It may be reasonable, however, to assume that an error profile for many of these measurements might be similar to those for similar measurements made in other surveys (see, for example, Brooks and Bailar's (1978) error profile for the CPS labor force measurements). Nonetheless, the fact that the population of interest is quite unlike a cross-section of the adult population would argue for caution in making such an assumption.

#### Time Period for Evaluation

If we assume that recall and other errors in the interview data (and processing and reporting errors in Social Security earnings records) are not too troubling, the CLMS data provide interview data for three years after a young person entered a program. The Social Security earnings records extend this time frame even further (e.g., for the 1975 program entrants we may have earnings in 1983).

For the NLS data we have waves of data covering five years of actual interviewing. Since the NLS obtained retrospective data in its initial interview, for some respondents we will have outcomes that were measured more than five years after participation in a program.

### General Issues of Method<sup>1</sup>

Despite the distinct advantages of the large and well-executed CLMS and NLS samples, there are several serious deficiencies in the data. First and foremost, neither data set is a true experimental sample which randomly sorts youths into participant and control groups. As a result, comparison groups must be constructed by a "matching" procedure using demographic variables and preprogram earnings. For the CLMS, a comparison group was drawn from Current Population Survey data and for NLS, a comparison group was drawn from nonparticipants within the sample itself.

However, for reasons discussed in greater detail in the final section of this appendix, these synthetic "control" groups are not entirely adequate. For the one empirical study for which there is explicit evidence, it was found that participants appeared to be more disadvantaged than nonparticipants in ways that could not be "matched" with the available data. As a result estimates of net program impact were downwardly biased.

Information on several key variables is missing in one or another of the samples. In particular, lack of location data in the CLMS makes it impossible to take account of variations across local labor markets or to assess the site-specific component of the variance in outcomes. Furthermore, the matching CPS file fails to record either participation in CETA or subsequent schooling. These deficiencies can lead to underestimates of the overall impact of CETA participation (note 1).

Finally, the analyses of these data frequently ignore the fact that both the NLS and the CLMS have complex sample designs. The variances for estimates derived from such designs can differ considerably from

---

<sup>1</sup>For the CLMS analyses, further problems are posed by deficiencies in the information available from the CPS files used in matching. First, many CPS youths will have been enrolled in CETA programs themselves, but neither the CPS nor the accompanying SSA files records such participation. SRI estimates these cumulative probabilities over the period of 1975 to 1978 to be: 12.2 percent for adult men, 14.9 percent for adult women, 31.1 percent for young men, and 30.7 percent for young women. Second, time spent in school is not recorded for the CPS sample during the postprogram period. As a result, net impacts can only be estimated for earnings. However, if CETA graduates are more likely than others to seek further education, this will also tend to result in lower earnings for them in the first and second postprogram years, at least. Not only will a negative earnings impact be accentuated, but there will be no record of what is considered an additional, positive impact of CETA, that is a return to school on the part of dropouts or other highly disadvantaged groups. The Urban Institute has found, for example, that negative findings for youths in classroom training may be associated with their increased time spent in school in postprogram years. Such data limitations thus convey an overly negative impression of the impact of the program.

those estimated using simple random sample (SRS) formulas (i.e., those produced by the statistical routines of the most widely used computer packages, e.g., SPSSX, SAS).

These issues are discussed in greater detail below.

#### FINDINGS OF STUDIES USING CLMS DATA BASE

Three studies compare CETA participants from the fiscal 1976 and fiscal 1977 Continuous Longitudinal Manpower Surveys with comparison groups selected from the March 1976 or March 1977 Current Population Surveys. These studies were conducted by Westat (1984), SRI International (1984), and the Urban Institute (Bassi et al., 1984). A fourth study conducted by Mathematica Policy Research (Dickinson et al., 1984) compares net-impact estimates derived from such CPS comparison groups with those from a true experimental comparison of participants and controls (using data from the Supported Work demonstration program).

The first three CLMS reports differ in their selection of CPS comparison groups and in the analytic models used for net-impact estimation. However, their findings show similar patterns: negative and statistically significant (or negligible) net impacts on post-CETA earnings for young men and positive, but generally insignificant net impacts for young women. Only on-the-job training seems to produce positive gains for most groups, and work experience is universally negative in its impact on earnings. Among the studies, the Urban Institute reported the greatest negative impacts and Westat the most positive; the SRI results were in-between. Again, however, it is important to note that these findings may be biased estimates of the true impacts of CETA on youths and, as such, may offer an inappropriate assessment of the overall program.

#### Analysis Strategies

Each of the three CLMS-CETA evaluations used three data sets: (1) CETA participants selected from the Continuous Longitudinal Manpower Survey of program entrants from July 1, 1975, to June 30, 1976, and/or from July 1, 1976, to June 30, 1977 (Westat also conducted some preliminary analysis on an earlier cohort); (2) nationally representative samples of individuals from the Current Population Survey of March 1976 and/or March 1977; and (3) earnings (up to the Social Security tax limit) for both CETA participants from the CLMS and individuals selected from the CPS. Net-impact estimates are based on differences in preprogram and postprogram earnings between the CLMS-CETA participants and earnings changes for similar individuals from the CPS file.

This combination of data appears to offer many distinct advantages over the YEDPA studies discussed in Chapters 4 through 8. The samples are large and nationally representative and cover several years of CETA programs. Annual earnings data are available from Social Security Administration (SSA) files from 1951 to 1979. Comparable data are

available on all major programs (classroom training, on-the-job training, Public Service Employment, and work experience) for the same years and for similar economic conditions. Finally, the CLMS file contains detailed data on participants from prime sponsor records, individual interviews at entrance into CETA, and subsequent interviews up to 36 months later. For these reasons and others, Westat has characterized these data as superior to any data set that has previously been available for evaluating large scale federally funded employment and training programs.

The problems of net-impact estimation, however, are substantial and arguments for one method or another have been a central focus of each of the three major studies of CETA using the CLMS data base. Because the disagreements are often sharply drawn and because they result in wide variations in net-impact estimates, these analytic issues are discussed briefly here. No attempt is made to resolve disputes in one direction or another, except insofar as the Supported Work evaluation provides evidence suggesting bias in all of the estimates (see final section of appendix).

Analytic and statistical problems fall basically into three main categories:

1. preliminary screening of individuals from the CLMS and CPS files on the basis of missing or nonmatching data, termination of program participation before 8 days, and similar factors. Such prematch deletions exclude as much as 30 percent of the original sample;
2. selection of a comparison sample from the CPS to match CETA participants along earnings-related dimensions; and
3. specification of a linear regression (or other) model of earnings (and accounting for "selection bias").

Basically, the three studies may be distinguished in the following ways: Westat devoted substantial resources over several years to creating a comparison file with a "cell matching" and weighting technique, but ultimately used a fairly straightforward regression analysis to estimate net impacts. Using these methods, net impacts for youths were generally found to be negligible for men and positive for women (few precise figures for youths were provided in their study). SRI, in a subsequent study, focused on an alternative method of selecting a comparison file using a Mahalanobis or "nearest-neighbor" matching technique, but also adopted straightforward regression analysis for most of its estimates, particularly of youths.

For all the attention paid by both Westat and SRI to the selection of the comparison group, SRI found that the two methods produced similar results, all else held equal. The more negative results presented as findings of the SRI study stem primarily from differences in the preliminary screening process and from updating of earnings with new information from SSA. (SRI's net-impact estimates are \$591 for young men and \$185, but statistically insignificant, for young women.) Finally, the Urban Institute adopted the Westat comparison file on youths but used a "fixed-effects" estimator to control for bias in

selecting participants into CETA and reported substantially more negative net impacts (a range of  $-\$515$  to  $-\$1,303$  for young men and  $-\$23$  to  $-\$391$ , but statistically insignificant, for young women).

It appears, therefore, that the primary differences in the net-impact estimates are based not in the time-consuming creation of comparison files, but in the preliminary screening of the CPS and CLMS files and in the specification of an earnings model. The Supported Work study by Mathematica provides similar evidence that the particular procedure used to select a comparison sample is less important than the net-impact estimation model (a fixed-effects estimator led to a more negative result than a linear model of postprogram earnings).

The basic goal in selecting a comparison file from the CPS is to find a group of individuals who closely resemble the CETA participants from the CLMS. Lacking a true (experimental) control group, a comparison group procedure is a next-best approach for comparing the earnings and employment outcomes of those who participate with those who do not. Net-impact estimates in these analyses are simply the coefficient on program participation in an earnings regression that controls for background characteristics and other earnings-related differences in a composite sample of youths from CLMS and CPS.<sup>2</sup>

Three basic techniques of selecting a comparison file have been used in these studies:

1. random sampling of CPS cases screened only for program eligibility;
2. stratified cell matching whereby a list of earnings-related variables is generated, CLMS participants are arrayed across cells by these variables, and CPS cases are matched and weighted to produce a similar distribution of participants and nonparticipants. Substantial "collapsing" of cells is required since the number of cells is large even for a small list of variables; and
3. statistical matching based on predicted values of earnings or the "nearest neighbor" technique of minimizing a distance function of a weighted sum of differences in earnings-related characteristics of the individual.

Several tests of the "success" of the CPS match are available. These are similarity in demographic or background characteristics

---

<sup>2</sup>If earnings functions could be correctly specified, a close matching of the CLMS and CPS data files would not be so important. But known nonlinearities, interactions of variables, and other complexities of labor market behavior across the population at large make impact estimates from a simple, linear and additive model highly suspect. Breaking down the files into subgroups (as by sex and race for the Urban Institute and by sex and program activity for Westat and SRI) would handle some matching of youths on other earnings-related characteristics and would also make net-impact estimation more precise for the range of people who are likely to enroll in CETA.

(especially those variables that are important determinants of earnings), similarity in preprogram earnings, and similarity in preprogram earnings functions. In particular, a test may be made of whether a CETA participation dummy variable is predictive of (or correlated with) a preprogram dip in earnings, as an indicator that program administrators may be "creaming" those individuals with a temporary drop in a relatively high "permanent" income stream.

A fixed-effects estimator is designed to control for such creaming and other sample selection bias by "differencing" a base-year and a postprogram year earnings equation. Any unobserved characteristics that lead to participation in CETA, but also affect earnings, are assumed to be constant over time and can be accounted for in such a procedure. If there is creaming based on a transitory preprogram drop in income, then the base-year must be chosen a year or two earlier to reflect a more permanent income trend.

In the majority of cases in the three reports, the CPS comparison groups pass the tests of similarity to CLMS/CETA participants. For example, as a result of cell matching or nearest-neighbor matching, the CPS pool is winnowed from a largely white sample of in-school youths or high school graduates from families above the poverty level to a mixed black/white sample that includes large numbers of high school dropouts from families below the poverty line. The comparison groups also resemble CETA participants in preprogram earnings. Matching on such background characteristics and preprogram earnings, of course, does not necessarily equalize unmeasured characteristics (e.g., actual or perceived motivation, ability)--a point to which we shall return.

#### Westat Findings

In 1980, Westat began to release a series of net-impact studies based on CLMS and CPS data. Comparison groups were created using stratified cell-matching techniques for CETA entrants in the first half of 1975, and for fiscal 1976 and 1977. Cells were defined by such variables as age, race, sex, family income, and education. Two basic matching subdivisions were made: one divided the CLMS sample into low, intermediate, and high earners and constructed separate CPS comparison file for each, and a second divided the program activities into: classroom training, on-the-job training, public service employment, work experience, and multiple activities. Because the latter match was more "successful" in terms of passing statistical tests of similarity between groups, it was used in most of the later Westat studies. Net impacts were estimated for three post-CETA years for the fiscal 1976 group and for two years for the fiscal 1977 group.

Westat's (1984) report summarizes their findings over the last several years. Although the report presents very few specific results for young men and women, its overall conclusions for adults are of interest. Key findings are the following:

- Statistically significant positive impacts for both cohorts and all postprogram years; estimates ranged from a low of \$129 per year to a high of \$677;



- Among programs, classroom training and on-the-job training show the highest net impacts and work experience the lowest; these rank orders are relatively stable across cohorts and postprogram years;
- For the first cohort, there was a marked difference in net impacts by sex--males experienced statistically insignificant gains and females experienced significant gains;
- For the second cohort, however, net impacts converged for men and women at statistically significant levels;
- Higher net impacts for low earners (less than \$2,000 in 1972 and 1973) than for high earners;
- Positive gains from "placement" in a job at termination and increasing gains with length of stay in the program;
- Substantially higher net impacts for the second cohort than for the first; these are attributed to a dramatic increase in net impacts for men, a decline in the proportion of youths with work experience, and across-the-board increases in all programs.

Specifically for youths, Westat found,

- Youth work experience programs are statistically insignificant for all cohorts and postprogram years.

Other specific youth-related findings are not reported in Westat (1984), but the Urban Institute has characterized Westat's results from earlier reports, as follows:

In looking at youth, Westat (1982) has found that for those youngsters 14 to 15 years old, CETA has had little overall impact. For other young workers net gains are found, being highest once again for OJT, followed by PSE and classroom training, and being negligible for work experience. The results found for young workers also tend to persist in the second postprogram year. Westat also produced a technical paper focusing on youth in CETA (1981) in which net gains were broken down by sex. As with adults, net gains were greatest for young females, being negligible or insignificant for males. After classifying youth according to their attachment to the labor force, net earnings gains were found to be greatest among structurally unemployed or discouraged workers.

#### SRI Findings

SRI's analysis differs from Westat's in two key respects: in the selection of the comparison group and in its "sampling frame." SRI's comparison groups were drawn by use of a "nearest-neighbor" matching procedure based on minimizing the "distance" of CLMS participants and selected CPS matches along earnings-related variables. SRI's sampling frame differed from Westat's in the following specific ways: development of calendar year cohorts rather than fiscal year cohorts; SRI

TABLE D.1 SRI Estimates of Net Impact of CETA on SSA Earnings (Standard Errors in Parentheses)

---

| Subgroup                 | SSA Earnings<br>(dollar impacts) |
|--------------------------|----------------------------------|
| Adult men<br>(N=6,144)   | -690<br>(139)                    |
| Adult women<br>(N=5,438) | 13<br>(116)                      |
| Young men<br>(N=3,298)   | -591<br>(167)                    |
| Young women<br>(N=2,826) | 185<br>(139)                     |

---

NOTE: Published standard errors for estimates appear in parentheses but are likely to be inaccurate; see note 5.

SOURCE: SRI International (1984).

inclusion (versus Westat exclusion) of individuals who received only "direct referrals" among those who received fewer than eight days of treatment; SRI exclusion (versus Westat inclusion) of individuals who worked in 1975 but were out of the labor force in March 1976; and use of a different set of rules for excluding individuals if key CPS or CLMS codes did not match their SSA codes.

SRI's model differed from Westat's only in the addition of several variables, such as veteran status and earlier earnings and the square of 1975 SSA earnings. (Table D.1 presents SRI estimates of net impacts of CETA on earnings for all participants in 1978.) SRI also experimented with fixed-effects estimators for adult men and women, but argued that they were not appropriate for youths just beginning work.

SRI's estimates of program effects were substantially below Westat's for both adults and youths, and the authors spent considerable time in identifying the sources of those differences. From their analyses, the SRI authors concluded that most of the differences could be attributed to choices made in the sampling frame and to an updating of 1979 SSA earnings.<sup>3</sup>

---

<sup>3</sup>Net impacts were minimally sensitive to the estimation model or to the matching technique used.

SRI (1984) reported the following findings for 1976 CETA enrollees:

- Participation in CETA results in significantly lower postprogram earnings for adult men (-\$690) and young men (-\$591) and statistically insignificant gains for adult women (+\$13) and young women (+\$185).
- All program activities have negative impacts for men, but adult women benefit from PSE and young women from OJT. Work experiences have negative impacts for all age and sex subgroups.
- Both male and female participants are more likely to be employed after CETA, but males are less likely to be in high-paying jobs or to work long hours.
- Length of stay in the program has a positive impact on postprogram earnings; turning points for young men are at 8 months and for young women at 1 month.
- Placement on leaving the program leads to positive earnings gains.

#### Urban Institute Findings

The Urban Institute used Westat's match groups from the CPS and estimated net impacts for six race/sex groups (male/female by white/black/Hispanic). Both random-effects estimators and fixed-effects estimators were used to identify net impacts, but the emphasis was on fixed-effects models which controlled for selection bias. Net impacts were estimated for two postprogram years, 1978 and 1979. (Table D.2 presents net impacts estimated in the Urban Institute analysis.)

The Urban Institute (Bassi, et al., 1984) found, for youths:

- Significant earnings losses for young men of all races and no significant impacts for young women; these impacts persist into the second postprogram year;
- Significant positive net impacts for young women, particularly minorities in Public Service Employment and on-the-job training and significant negative or insignificant net impacts for all groups in work experience;
- Among subgroups, the most negative findings were for white males, the most positive for minority females;
- Older youths (22-year-olds) and those who had worked less than quarter time had stronger gains or smaller losses than the younger group or those who had worked quarter time or more;
- Earnings gains resulted primarily from increased time in the labor force, time employed, and hours worked, rather than from increased average hourly wages.

#### FINDINGS OF STUDIES USING NLS DATA BASE

Two major studies have used the National Longitudinal Survey data base to estimate the aggregate effects of government-sponsored employ-

TABLE D.2 Urban Institute Estimates of Net Impact of CETA Participation on Earnings of Youths Under Age 23, for Three Models

| Race/Sex Group   | 1978            |                    |                    | 1979            |                    |                    |
|------------------|-----------------|--------------------|--------------------|-----------------|--------------------|--------------------|
|                  | RE <sup>a</sup> | FE-75 <sup>a</sup> | FE-74 <sup>a</sup> | RE <sup>a</sup> | FE-75 <sup>a</sup> | FE-74 <sup>a</sup> |
| White females    | - 109<br>(.930) | - 23<br>(.194)     | - 68<br>(0.57)     | - 190<br>(1.33) | - 92<br>(0.62)     | - 136<br>(0.92)    |
| Black females    | - 201<br>(0.89) | - 77<br>(0.34)     | - 159<br>(0.71)    | 24<br>(0.09)    | 170<br>(0.61)      | 88<br>(0.32)       |
| Hispanic females | - 391<br>(1.10) | - 261<br>(.716)    | - 231<br>(0.64)    | - 501<br>(1.14) | - 348<br>(0.79)    | - 318<br>(0.71)    |
| White males      | - 593<br>(4.23) | - 515<br>(3.73)    | - 576<br>(4.12)    | - 795<br>(4.47) | - 728<br>(4.11)    | - 789<br>(4.43)    |
| Black males      | - 989<br>(3.67) | - 681<br>(2.57)    | - 758<br>(2.85)    | -1167<br>(3.42) | - 748<br>(2.16)    | - 822<br>(2.38)    |
| Hispanic males   | -1155<br>(2.56) | -1193<br>(2.61)    | -1303<br>(2.82)    | -1009<br>(1.78) | -1080<br>(1.89)    | -1190<br>(2.07)    |

NOTES: Coefficients represent dollar impacts on social security earnings. Published t statistics are in parentheses, but are likely to be inaccurate; see note 5.

<sup>a</sup>Models: RE--random effects; FE-75--fixed effects, base period 1975; and FE-74--fixed effects, base period 1974.

SOURCE: Bassi et al. (1984).

ment and training programs on youths. One study (Moeller et al., 1983) was conducted by the Policy Research Group (PRG) of Washington, D.C., and the other (Hahn and Lerman, 1983) by the Center for Employment and Income Studies (CEIS) of Brandeis University. Both studies evaluated the effects of CETA programs on youths, although the PRG study expanded its scope to include schooling programs, such as vocational education.

The estimates made by both studies indicate relatively modest effects of employment and training programs on the subsequent income, employment status, and educational attainment of the youths who participated in those programs. For CETA programs both studies find negative overall effects of CETA on employment, although PRG reports some positive effects at two years after CETA completion. Reviewing PRG results and their own findings, the CEIS authors (Hahn and Lerman, 1983:84) dourly conclude an appendix to their chapter entitled "Did the CETA System Work for Disadvantaged Youth?" by noting:

To conclude, both the PRG results and our own show negative and significant effects of CETA on employment variables. It is only after going out two years in time after CETA completion that the PRG report finds evidence of a positive, significant effect and that on only one variable, unsubsidized earnings. We cannot confirm this positive effect, but it would not be

inconsistent with our results. It is difficult to claim this as an impressive success for CETA. [emphasis added]

Since the substantive findings from the NLS analyses are generally consistent with the weak and generally negative findings from the CLMS analyses we do not review them in great detail. Instead, we briefly describe the analysis strategies and findings of each of the studies then turn to a consideration of the potential for bias introduced by use of statistical matching procedures rather than random assignment to construct "control" groups in the CLMS and NLS analyses.

#### CEIS Procedures and Findings

The analysis reported by Hahn and Lerman (1983) employs a nearest neighbor matching procedure<sup>4</sup> to construct a "matched" control group to be compared with the 1,114 respondents in the NLS youth sample who reported participating in CETA programs. (The matched sample was constructed by selecting respondents from among the 4,608 NLS respondents who reported that they had not participated in a CETA program and who were neither in the military nor had family incomes above \$25,000 in 1978, and who did respond to the questions used to construct the eight matching variables.)

The variables used for matching were: sex, race, age, family size, family income (in 1978), weeks employed (in 1978), whether the youth was living at home, and whether the youth was a high school dropout. All the matching variables were derived from the 1979 survey. The resulting "matched" sample was then used to estimate the impact of (prior) participation in CETA programs on earnings and employment in 1979 and later years. As with the CLMS matchings, the CEIS analysis of the NLS data base takes a pool of largely white, middle-class youths and produces a "matched" sample which is 65 percent black and has a mean family income of \$8,790 (in 1978).

The CEIS analysis concentrates initially on employment in unsubsidized jobs as its major outcome measure. The authors argue that this is the appropriate outcome measurement for initial study since "politically the motivating concern in establishing a CETA program" was to increase the likelihood that disadvantaged youths could find employment in the regular (i.e., non-CETA) labor market. Their key analysis derives an estimate of program effects for a regression model that incorporates 20 other independent variables.

The coefficients estimated in this analysis are reproduced in Table D.3 together with the published t-ratios. (The latter statistics appear to ignore the complex sampling design used in the NLS and thus

---

<sup>4</sup>Hahn and Lerman (1983:75) used a "nearest available Mahalanobis metric matching" method suggested by Rubin (1979).

are likely to be inaccurate.<sup>5</sup>) It will be seen from Table D.3 that the net-impact estimates derived from this analysis are quite negative,  
(text is continued on p. 269)

<sup>5</sup>A complete assessment of the variance and bias components of estimates derived from surveys of complex design requires an assessment of effects arising from sample design, interviewer contributions to the variance, and the effects of sampling frame bias and nonresponse, among other factors. These variance components in survey estimates are typically referred to as "design effects."

It is often the case that the secondary analyst lacks sufficient information (coded into the sample data) to make adequate assessments of such components of the variation in his or her estimates. In the NLS (and many other surveys conducted by quality-conscious research organizations), information is available concerning the sample design used in the survey (see Frankel et al., 1982).

Like many large-scale household surveys, the NLS used a multistage area probability sample with sample clustering at the block level. For reasons of cost and efficiency, self-weighting simple random samples (SRS) are seldom used in such large-scale surveys. Rather, sample designs incorporate a multi-stage selection process and some type of sample clustering. Instead of a random draw of individuals, small geographical areas (e.g., blocks in cities) are sampled, and then a number of individuals in that area are selected for interview. This group of individuals is referred to as a sample cluster. Cluster sizes of 5 to 10 individuals per "block" are common in national surveys. In personal interview surveys, such sample clustering is a practical necessity, since any random draw of subjects in a national survey might require some interviewers to travel hundreds of miles between successive interviews.

Because a sample of, say 5,000, may thus consist of 500 randomly selected "blocks," from which 10 individuals were selected, the variances of both univariate and multivariate statistics are not accurately portrayed by the well-known formulas that apply to simple random samples. This is so because the sampling of individuals was not independent of their location. To the extent that individuals within a selected location are homogeneous with respect to a given measurement, the effect of the use of a clustered sample is to increase the variance of the estimates (above those that obtain for SRS designs)--or, to express it another way, to decrease the effective sample size from the number of individuals ( $N_i$ ) toward the number of locations that were chosen ( $N_c$ ). In extreme cases, e.g., the use of such samples to estimate the distribution of the population across (self-perceived) rural versus nonrural areas, the effective sample size may approach  $N_c$  since almost all of the extant variation occurs between rather than within sample clusters--and hence we have a sample of  $N_c$  clusters, where almost all of the  $N_i$  respondents in any one cluster give the same response when asked whether they live in a rural area. Another way of expressing this result is to say that the intra-cluster correlation of response approaches 1.0 for this variable. For more rigorous theoretical treatments of this matter, see Hansen et al.

(Footnote 5 is continued on pp. 268-269)

---

(1953) or Kish (1965); for an elementary introduction, see Blalock (1972:520-530).

Of course, residence in a rural area is an extreme case, but high intracluster correlations arise for many other variables of interest. The type of dwelling unit (apartment versus house) has similar characteristics, and even such crucial variables as race and income show relatively high intra-cluster correlations in sample surveys as a result of the residential segregation of neighborhoods. (Persons of the same race, income, and so on tend to live together.)

This has important practical implications for the survey analyst since the widely used algorithms for computing variances for simple random samples (e.g., those of the SPSSX, SAS, and BMD computer packages) are not appropriate. Use of SRS formulas to calculate the sampling variances from commonly used cluster designs may considerably understate standard errors for means, regression coefficients, and so on. Since variances and their roots figure crucially in all inferential statistics, serious errors of inference can be made.

Design effects (deff) that arise from the use of complex sample designs are commonly expressed (see Kish, 1965:258ff) as the ratio of the actual variance of a sample estimate to the variance of an SRS sample of equal size. The square root of the deff ratio (which is referred to as deft) can be used to "deflate" statistics derived under SRS assumptions. For example, in the case of the t-statistic:

$$\text{deft} = (\text{deff})^{0.5}$$
$$t_{\text{true}} = t_{\text{srs}}/\text{deft}$$

A careful study (Frankel, 1971) of design effects for eight variables in the Current Population Survey found values of deft that ranged from 1.1 to 1.5. (The variables studied were number of persons in household, number aged 0-17, number in labor force, household income, and, for the "head of household," income, age, sex, and educational attainment.) In discussing the practical importance of such studies, Frankel (1971:1-2) observed:

As social scientists become more mathematically sophisticated and attempt to use sample survey data to uncover multivariate relationships, the gap between the assumptions of existing statistical theories and the actuality of sample designs used to collect data, make the valid use of standard inferential techniques tenuous. Naive researchers may be unaware of this gap and may use inference which assumes that their sample is a simple random selection of elements. However, more sophisticated social scientists are faced with the task of making statements like that of Blau and Duncan in their study, The American Occupational Structure: "We do not know, however, how other statistics, such as regression coefficients . . . and F-ratios, are affected by the departure of the sampling design from simple random sampling. Only very rough guesses about standard errors can be made. . . ."

e.g., -\$1,640 in 1980 earnings (from unsubsidized employment). The CEIS study further examines some of these outcomes for particular population subgroups and youth participating exclusively in "skills training" programs. With a few modest exceptions, however, these analyses lead the authors to similar conclusions, i.e., no (positive) net impact of CETA participation.

While these analyses share many characteristics with the CLMS analyses (and share the potential for "matching bias" that is discussed in the final section of this appendix), they do have unique aspects that should be noted. A critical one is their reliance on respondents' self-reports of whether they participated in CETA.<sup>6</sup> Although no estimates of the error and bias components of these measurements are proffered, it is unlikely that these crucial variables will be reported without error. Also, in contrast to the SSA records used in the CLMS analyses, the CEIS study must rely on the self-reports of earnings in the NLS--and indeed must then use other self-reports to characterize those earnings as "subsidized" or "unsubsidized." Here, again, the potential for inaccuracies in such reports (involving both random and systematic errors) could potentially affect the outcomes obtained in the CEIS analysis.

---

Frankel's (1971) findings suggest that standard errors for his eight variables would be understated by about 10 to 50 percent if SRS formulas were used (with the larger numbers applying to multiple correlation statistics, the smaller to multiple regression coefficients, and means and simple correlations resting in the middle). He has, however, no available data for race, which is quite highly clustered geographically (racial discrimination and other factors produced a very nonrandom distribution of blacks and whites across "blocks" and neighborhoods in the United States). Statistical research on health surveys (Landis et al., 1982) indicates that defts in excess of 2.0 occur for such things as the partial regression coefficient representing the "effect" of race on health conditions (e.g., standard errors for the "effects" of race on number of decayed teeth, net of age, sex, and consumption of sugar, may be underestimated by a factor of 2 or more if SRS formulas are used). It is difficult to assess what analogous values might be in earnings equations, but caution is clearly advisable in interpreting the estimates provided by the CEIS and PRG analyses.

<sup>6</sup>CLMS, in contrast, sampled participants from lists using administrative records (from prime sponsors). The CPS match samples used in the CLMS studies are, however, subject to similar but even more severe concerns. It is not possible to identify "program participants" in the CPS match sample, and thus the CLMS analyses must contrast a sample of individuals who all participated in CETA with a cross-sectional sample containing both participants and nonparticipants.



**TABLE D.3 CEIS Estimates of Net Impact of Participation in CETA on Unsubsidized Net Earnings (in dollars per year)**

| Independent Variables                      | Unsubsidized Earnings |                    |
|--|-----------------------|--------------------|
|  | 1979                  | 1980               |
| Participation in CETA                      | -675.9<br>(-3.92)     | -1640.2<br>(-7.54) |
| Reservation wage                           | 661.4<br>(5.4)        | 980.8<br>(6.28)    |
| Has child                                  | 142.7<br>(0.45)       | 130.3<br>(0.31)    |
| White                                      | -74.7<br>(-0.37)      | 17.3<br>(.07)      |
| Has illicit income                         | 9.5<br>(.04)          | -57.8<br>(0.20)    |
| Rotter scale<br>(locus of control)         | 65.7<br>(0.90)        | -70.2<br>(-0.70)   |
| Family income 1978                         | .047<br>(3.56)        | 0.060<br>(3.48)    |
| Area unemployment<br>rate 1979             | -22.0<br>(-0.61)      | 27.14<br>(0.60)    |
| Weeks employed 1978                        | 57.2<br>(10.97)       | 49.85<br>(7.43)    |
| High school dropout<br>1979 interview date | 234.7<br>(0.92)       | -525.0<br>(-1.59)  |
| Female                                     | -867.3<br>(-4.51)     | -1077.4<br>(-4.32) |
| Family size                                | -148.0<br>(-3.49)     | -160.7<br>(-2.88)  |
| Knowledge-of-world-of-<br>work scale       | 11.8<br>(0.23)        | 14.2<br>(0.21)     |
| Numerical operations<br>standard score     | 88.0<br>(0.77)        | 55.4<br>(0.37)     |
| Age at 1979 interview                      | 287.8<br>(4.44)       | 407.9<br>(4.81)    |
| Mechanical comprehension                   | -79.8<br>(-0.56)      | -140.9<br>(-0.78)  |
| Does not live at home                      | -23.1<br>(-0.08)      | -537.4<br>(-1.44)  |
| Paragraph comprehension<br>standard score  | 137.1<br>(0.95)       | -95.7<br>(-.50)    |
| Math knowledge standard<br>score           | -242.3<br>(-1.48)     | -327.2<br>(-1.55)  |
| Word knowledge standard<br>score           | -98.7<br>(-0.62)      | -139.0<br>(-0.68)  |
| Arithmetic comprehension<br>scale          | 250.3<br>(1.34)       | 368.8<br>(1.54)    |
| Constant                                   | -3908.8<br>(-3.38)    | -4456.9<br>(-2.92) |
| R <sup>2</sup>                             | .24                   | .22                |
| N  | 1266                  | 1120               |

NOTE: Published t statistics appear in parentheses beneath coefficients, but for reasons discussed in note 5, these values are likely to be inaccurate.

SOURCE: Hahn and Lerman (1983).

### PRG Procedures and Findings

While the PRG analyses differ in their details from those of CEIS, the basic strategy was the same. PRG used the NLS youth sample to (1) identify all participants in government employment and training programs, (2) construct a comparison group of nonparticipants, and (3) estimate a model for the outcome variables of interest. The PRG analysis of the NLS data base differs from CEIS's in its use of a wider range of outcome measures (including earnings, employment, educational, and marital outcomes) and a somewhat different strategy for constructing a comparison group of nonparticipants.

PRG used what it described as a "stratified random sampling procedure" to select the comparison group, but the description of this procedure is unclear in some respects. The authors' (Moeller et al., 1983:E-1) "overview" of the procedure is stated as follows:<sup>7</sup>

Both the CEIS (1982) and Westat (1980) studies . . . adopted a "match" procedure for selecting a control group (hereafter referred to as the CGRP). We instead chose to use a stratified random sampling procedure for its computational advantages and sound statistical approach to selecting the CGRP sample. In combination with a reasonably complete control variable specification in the outcome regressions, weights for the two samples to equate the number of participant and comparison group members within each stratification cell, and a selectivity bias correction for "unmeasured" differences between the participant and CGRP members, we did not judge the additional computational burden of a match procedure to be warranted.

It appears that this procedure involved the construction of a synthetic variable representing the socioeconomic status (SES) of the respondent and then cross-classifying participants and nonparticipants by SES, sex, race, local unemployment rate, region. Prior to the cross-classification, each of these variables was dichotomized (e.g., local unemployment: 0-5 percent versus 6 percent or more), except for region, which had four categories. Nonparticipants were then selected at random from within the resulting 128 cells with the probability of

---

<sup>7</sup>This text is an accurate reproduction of the PRG statement (the original is also garbled). It should be noted, too, that the "selectivity bias correction" analysis was not included in the PRG report, and according to statements elsewhere in the report these analyses were not performed.

selection for each cell being equal to the proportion of the participant sample that fell into the same cell.<sup>8</sup>

Two aspects of the PRG analysis are troubling. First, the authors used an ordinary least squares procedure to estimate their model equations where some of their dependent variables take only two values (e.g., 0: out of school, 1: in school). This, in addition to use of procedures that assume simple random sampling, raises doubts about the accuracy of the reported significance levels. On a more substantive level, we note that the authors never combine their separate analyses of employment status and education, and so we cannot tell to what extent the decreased earnings of CETA participants might be due to the increased enrollments in school. If this were to account for an important share of the observed income drop, one might characterize the earnings decline as an investment of foregone earnings in education rather than a negative outcome of CETA.

Additional PRG analyses estimate that CETA had few impacts on other outcomes (e.g., receipt of welfare or unemployment income, criminal behavior, graduation from high school, disciplinary problems in school, or health status) that were reliably different from zero, based on the

---

<sup>8</sup>Because the published description of these procedures is garbled in places, it is not entirely clear how this selection strategy would differ from a cell-matching procedure--except for the arbitrary manner in which the size of the "control" sample is set (i.e., by specifying a sampling fraction). In other details, there are also several puzzling aspects. For example, great efforts are put into constructing a composite family income and social status indicator (from a regression using 97 variables reflecting aspects of youths' income and social status), but the resultant continuous variable (scaled in a metric of "expected" family income) is merely dichotomized (less than \$15,000 versus more than \$15,000 per year).

The resulting samples were then used to estimate program impacts by embedding a dichotomous program participation variable in equations predicting each of the outcomes shown in Table D.4. (Other independent variables in these equations were intended to control for region, age, race, pre-enrollment employment status, family income, marital status, educational level, and health status.) It will be seen from Table D.4 that across all time periods studied, PRG estimates that the net impact of CETA was -\$28 per month on earnings from unsubsidized employment. Estimated net impacts for other outcome variables are also negative or "insignificant." (Note, however, that the t-ratios are likely to be inaccurate since the PRG analysis treated the NLS data as if they had been derived from a simple random sample of the population; see note 5.) The sole positive result shown in this analysis is for education, for which it is estimated that the net impact of CETA was to increase the probability that the youth would remain in (or return to) school by 5.6 percent.

TABLE D.4 PRG Estimates of Net Impact of Participation in CETA on Employment, Earnings, Education, and Marital Behavior

---

| Outcome  | Impact of CETA    |
|--|-------------------|
| Months of unsubsidized employment  | -.051<br>(2.85)   |
| Unsubsidized earnings  | -27.698<br>(2.34) |
| Hours of unsubsidized employment per month <sup>a</sup>                  | -8.844<br>(3.89)  |
| Hours of unsubsidized employment with wages set by collective bargaining | -.008<br>(.92)    |
| Probability of being employed in unsubsidized job                        | .028<br>(1.59)    |
| Months of regular school   | .014<br>(.98)     |
| Probability of being in regular school                                   | .056<br>(3.45)    |
| Probability of being married   | -.088<br>(.54)    |

---

NOTE: Averages calculated over the youth's postprogram quarters, up to 12. The t statistics appear in parentheses, but are likely to be inaccurate; see note 5.

<sup>a</sup>This entry is listed in source as "months of unsubsidized employment," not hours, but this appears to be a typographical error, since it duplicates first entry in table.

SOURCE: Moeller et al. (1983).

PRG computations. The two exceptions were increased use of drugs among CETA participants (net impact +7.3 percent) and increased likelihood of being married (10.2 percent).<sup>9</sup> However, teenage matrimony would be unlikely to qualify as a positive outcome of a CETA program, and of

---

<sup>9</sup>Note that these two dependent variables are also dichotomies which were analyzed using OLS procedures.

course, the increased "use or sale of marijuana, hashish, or hard drugs" would be thought by most observers to be a negative social outcome. The sole optimistic findings of the PRG analysis occur two years after program completion. For selected quarters, the authors find evidence of positive net impacts of CETA on unsubsidized earnings and employment status. These impacts were not, however, reliably different from zero (using the authors' statistics) at the time of the last NLS measurements (33-36 months after program completion).

#### BIASES IN ESTIMATES OF PROGRAM EFFECTIVENESS ARISING FROM USE OF MATCHED SAMPLES RATHER THAN RANDOM ASSIGNMENT

Across the three CLMS studies, there is a pattern of preponderantly negative net impacts on youths, and the NLS studies show extremely weak effects of program participation. These results invite the conclusion that federally funded employment and training programs have had (in the aggregate) either little effect or a deleterious effect on the future earnings and employment prospects of the youth who participated in the programs. There is, however, reason to suspect (and empirical evidence to support the suspicion) that the foregoing estimates may be biased downward.

The reason<sup>10</sup> for this suspicion is that (despite intensive and varying efforts to select comparison groups similar to participants in youth programs and to control for selection bias through use of fixed-effects estimators) there may still be persistent and systematic, but unobserved, differences in the earnings profiles of comparison groups and true controls. Lower earnings, for example, might be due to such unobserved factors as (perceived or actual) differences in social attitudes, motivation, or ability between program participants versus a more "mainstream" comparison group.

A study by Mathematica (1984) provides important evidence on the potential for bias in the use of matching strategies such as those employed in the NLS and CLMS analyses reviewed above. The Mathematica study used data from a true experimental design that randomly assigned

---

<sup>10</sup>In addition to the potential bias in the matched control groups, there are two other reasons to question negative conclusions from the CLMS studies. The CPS lacks data on enrollment in CETA on the part of the comparison group and, as a result, positive net impacts may be underestimated since some of the "controls" were actually program participants. In addition, postprogram earnings are taken from SSA files, which contain no information on subsequent education or training. However, to the extent that CETA encourages further schooling, it reduces immediate postprogram earnings (and therefore lowers the net-impact estimate), but it probably should be viewed as a positive impact in its own right. Nevertheless, this interaction has not been and cannot be examined with the available data.

youths to be either program participants or controls. It then compared net impact estimates derived using the experimental design with estimates derived using the same sample of program participants but substituting various "matched samples" constructed from the CPS.<sup>11</sup> Mathematica examined net impacts based on simple differences in earnings gains, on a straightforward earnings regression model, and on a fixed-effects estimation model. Separate comparisons were performed for youths and women receiving Aid to Families with Dependent Children (AFDC).

Based on a true control group, Mathematica found in-program earnings gains and negligible postprogram effects for youths. Comparison of Supported Work participants and the CPS matched sample, however, yielded either insignificant or significantly negative effects. Moreover, the bias apparent in the match sample estimates was even greater using a fixed-effects estimator rather than a basic earnings model.

Figure D.1 (from Mathematica, 1984:Figure III.3) illustrates how this bias in the matched samples occurs. The age-earnings profiles of participants and true controls are dramatically different in the years following the program from the profiles of matched controls derived from the CPS (regardless of which of the three matching strategies is used).

While cell matching or statistical matching reduces mean differences in preprogram earnings and in background characteristics, subsequent earnings still diverge, for reasons that are left unobserved and unexplained, but which may have to do with actual or perceived differences in motivation, ability, or social attitudes (among other possible factors). (Alternatively, it may be the case that the scale of subsidized youth programs in 1978-1981 was sufficiently large that the programs indirectly improved the comparison groups' employment prospects. By temporarily withdrawing many participants from the competitive labor market for low-income youths, the programs may have enabled some nonparticipants to obtain more readily whatever unsubsidized jobs were available, and to this extent they boosted employment outcomes above what they would have been in the absence of such federally funded programs.)

Results for AFDC women provide an interesting contrast. In some instances, the Mathematica analysis finds an upward bias in estimates of program effects. But, in general, both the true control group analyses and the matched control group analyses show large and significant impacts both during and after the program. No clear pattern of

---

<sup>11</sup>Three techniques of matching were used: general eligibility screens, such as high school dropout; cell matching and weighting (similar to the technique used by Westat); and statistical matching based on predicted earnings (rather than on earnings-related variables, as done by SRI).

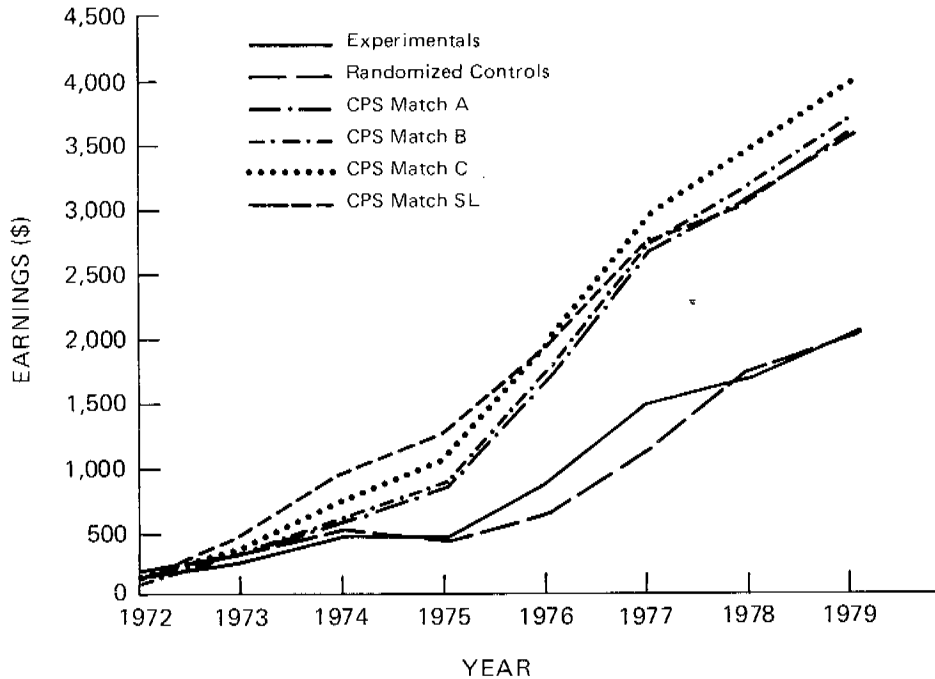


FIGURE D.1 Comparison of average SSA earnings for program participants and randomly assigned controls in Supported Work experiment to SSA earnings for "match groups" constructed from CPS sample using alternative matching strategies.

SOURCE: Mathematica Policy Research (1984:Figure III.3).

difference is found between the results obtained using a basic earnings model and a fixed-effects model. Mathematica argues that a similar negative bias probably exists for other CETA evaluations using constructed comparison groups rather than true controls, at least for youths, and it specifically cites the Westat, SRI, and Urban Institute findings in this regard.

Table D.5 (from Mathematica, 1984:Table IV.7) shows net impact estimates derived from Mathematica's analyses of Supported Work, together with estimates of overall program impact from the studies by Westat, SRI, and the Urban Institute. Mathematica acknowledges that its Supported Work sample is more severely disadvantaged and therefore more likely to have lower earnings profiles than the typical CETA youth participant. Nevertheless, there is some overlap of the two groups, and the Supported Work program did primarily provide supervised employment, which is an element of youth programs common to on-the-job-training projects, work experience projects, and public sector employment projects.

TABLE D.5 Alternative Estimates of Net Impact on Earnings (in dollars per year) of Participation in Supported Work and CETA Using Alternative Comparison Group Methodologies and Estimation Techniques

| Study and Methodology                               | Participant Group    |                               |
|---|----------------------|-------------------------------|
|   | Youths               | AFDC Women                    |
| Supported Work participants                         |                      |                               |
| Control group methodology                           | - 18                 | 351*                          |
| Comparison group methodology <sup>a</sup>           | -339 to -1179**      | 257 to 806**                  |
| CETA participants with comparison group methodology |                      |                               |
| Westat <sup>b</sup>                                 | --                   | 500** to 600**                |
| SRI using Westat comparison group <sup>c</sup>      | -122 <sup>d</sup>    | 488***                        |
| SRI using SRI comparison group <sup>c</sup>         | -524*** <sup>d</sup> | 246*                          |
| Urban Institute <sup>e</sup>                        | -515** to -1303**    | 556*** to 949*** <sup>f</sup> |

NOTES: Earnings are for 1978 for Supported Work and 1979 for CETA. Supported Work participants tended to enroll in the program slightly later than did the CETA participants included in the CETA net-impact studies. For this reason, 1979 outcome measures for the Supported Work samples are most nearly comparable to the 1978 outcomes for the CETA participant group studied. Published significance levels are denoted by asterisks, as follows:

- \* p less than .10
- \*\* p less than .05
- \*\*\* p less than .01

However, for reasons discussed in note 5, these levels may be inaccurate.

<sup>a</sup>Excludes results based on the random CPS samples meeting the Stage 1 screens.

<sup>b</sup>See Westat, Inc. (1980:Tables 3-6).

<sup>c</sup>See Dickinson et al. (1984, draft:Table V.3). Results reported pertain to enrollees during the first half of 1976.

<sup>d</sup>These figures pertain to male youths only. Data in the report did not permit the calculation of an overall impact for all youths. However, only 12 percent of the Supported Work youth were female.

<sup>e</sup>See Bassi et al. (1984:Tables 3 and 22).

<sup>f</sup>These figures pertain to female welfare recipients. Similarly large positive impacts were also estimated for all economically disadvantaged women.

Because of such similarities, Mathematica analysts argue that similar biases in estimates of program effectiveness may exist in the net impacts estimated by Westat, SRI, and the Urban Institute, and they conclude that "It is not possible to generate reliable net program impact estimates using ex-post comparison group procedures."



## CONCLUSION

While argument may be had (at great length given the dearth of reliable evidence) concerning the extent to which the Mathematica demonstration of bias in the matched sample methodology can be generalized, the study does highlight two separate problems in net-impact estimations using a matched comparison group:

1. the extent to which employment programs recruit or attract participants who differ from eligible nonparticipants in ways that may affect subsequent earnings; and

2. the extent to which such differences can be detected and controlled using available demographic or preprogram earnings data. For the latter problem youths presents a particularly difficult case for any match strategy because preprogram earnings data are either not extant or not reliable indicators of the uncontrolled variables that are of interest to program evaluators.<sup>12</sup>

Estimates of the magnitude and direction of the bias in matched-group evaluations are only available for the one youth program (Supported Work) whose experimental data were reanalyzed by Mathematica. From this reanalysis we have an elegant demonstration that commonly used "match" strategies would have yielded an inappropriately negative evaluation (where the experimental data indicate that the program had a null impact). There is an obvious temptation to leap from this one result to the assumption that biases equal in magnitude and direction affect all other "match group" studies. The available evidence, however, is not sufficient to warrant such a sweeping generalization. Until the methodological point is clarified by expanding on the provocative paradigm provided by the Mathematica analysts, there is considerable uncertainty as to the extent to which this finding will generalize to other program evaluations involving different populations of youths. Providing the requisite data will take a renewed commitment to conducting the randomized experiments needed to make estimates of the magnitude and direction of the biases involved in common matching strategies.

---

<sup>12</sup>In contrast, for adult women receiving Aid to Families with Dependent Children, it is apparently possible to control for such differences. Welfare payments are known, and preprogram earnings are a much better indicator for adults than they are for youths, and they can be used both in selecting a matched comparison sample and as a control variable in the net-impact estimation. Finally, the trend in preprogram earnings can be used to test for "creaming" or other sample selection biases that can be removed from the estimates.

REFERENCES

- Bassi, L.J., M.C. Simms, L.C. Burbridge, and C.L. Betsey  
1984 Measuring the Effect of CETA on Youth and the Economically Disadvantaged. Washington, D.C.: The Urban Institute.
- Blalock, H.  
1972 Social Statistics. New York: McGraw Hill.
- Blau, P., and O.D. Duncan  
1967 American Occupational Structure. New York: Wiley.
- Brooks, C., and B.A. Bailer  
1978 An Error Profile: Employment as Measured by the Current Population Survey. Statistical Policy Working Paper No. 3. Washington, D.C.: Office of Federal Statistical Policy and Standards.
- Center for Human Resource Research  
1982 National Longitudinal Surveys Handbook: 1982. Center for Human Resource Research. Columbus: Ohio State University.
- Dickinson, K.P., T.R. Johnson, and R.W. West  
1984 An Analysis of the Impact of CETA programs on Participants' Earnings. Menlo Park, Calif.: SRI International.
- Frankel, M.R.  
1971 Inference from Survey Samples: An Empirical Investigation. Ann Arbor, Mich.: Survey Research Center.
- Frankel, M.R., H.A. McWilliams, and B.D. Spencer  
1971 National Longitudinal Survey of Labor Force Behavior, Youth Survey: Technical Sampling Report. Chicago, Ill.: National Opinion Research Center.
- Hahn, A., and R. Lerman  
1983 The CETA Youth Employment Record: Representative Findings on the Effectiveness of Federal Strategies for Assisting Disadvantaged Youth. Final Report to U.S. Department of Labor. Center for Employment and Income Studies. Waltham, Mass.: Brandeis University.
- Hansen, M.H., W.N. Hurwitz, and W.G. Madow  
1953 Sample Survey Methods and Theory. 2 Vols. New York: Wiley.
- Kish, L.  
1965 Survey Sampling. New York: Wiley.
- Landis, J.R., and others  
1982 A statistical methodology for analyzing data from a complex survey. Vital and Health Statistics 2(92).
- Mathematica Policy Research  
1984 An Assessment of Alternative Comparison Group Methodologies for Evaluating Employment and Training Programs. Princeton, N.J.: Mathematica Policy Research.
- Moeller, J., R. Hayes, and I. Witt  
1983 Socioeconomic Impacts of Recent Government-Subsidized Employment and Training Programs on Youth. Washington, D.C.: Policy Research Group.
- Rubin, D.B.  
1979 Using multivariate match sampling and regression adjustment to control bias in observational studies. Journal of the American Statistical Association 74:318-328.

SRI International

1984 Analysis of Impact of CETA Programs on Participants Earnings.  
Stanford, Calif.: SRI International.

Westat, Inc.

1980 Continuous Longitudinal Manpower Survey. Net Impact Report No. 1. Impact on 1977 Earnings of New FY 1976 CETA enrollees in Selected Program Activities. Draft. Rockville, Md.: Westat, Inc.

1982 Continuous Longitudinal Manpower Survey. Net Impact Report No. 2. Impact of CETA on 1978 Earnings: Participants Who Entered CETA During July 1976 Through June 1977. Draft. Rockville, Md.: Westat, Inc.

1984 Summary of Net Impact Results. Rockville, Md.: Westat, Inc.