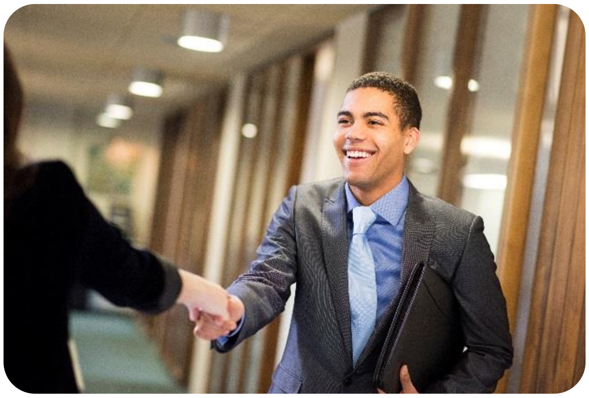


Career Pathways Long-Term Outcomes Study

Appendices for PACE Six-Year Impact Reports

OPRE Report 2022-69

March 2022

PACE

Pathways for Advancing Careers and Education

Appendices for PACE Six-Year Impact Reports

A Pathways for Advancing Careers and Education (PACE) / Career Pathways Long-Term Outcomes Study Publication

OPRE Report 2022-69

March 2022

David Judkins, Emily Roessel, and Gabriel Durham, Abt Associates

Submitted to:
Nicole Constance and Amelia Popham, Project Officers
Office of Planning, Research, and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

Contract Number: HHSP233201500069I,Task Order HHSP23337007T

Project Director: Larry Buron
Abt Associates Inc.
6130 Executive Boulevard
Rockville, MD 20852

 Suggested citation: Judkins, David, Emily Roessel, and Gabriel Durham. 2022. *Career Pathways Long-Term Outcomes Study: Appendices for PACE Six-Year Impact Reports,* OPRE Report 2022-69. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Contents

List of Exhibits

List of Boxes

Foreword

This volume provides details on the common methodology used to report on the six-year impacts of nine programs being evaluated as part of the **Pathways for Advancing Careers and Education (PACE) project**.

PACE evaluates the effectiveness of nine distinct programs, with study enrollment beginning in 2011. These programs were selected as potentially high-quality examples of programs that include key features of a career pathways framework. Each of these programs provided education or training to low-income adults for occupations expected to pay well and be in high demand locally.

Funded by the Office of Planning, Research, and Evaluation (OPRE) within the U.S. Department of Health and Human Services, Administration for Children and Families (ACF), the PACE project implemented the first large-scale, multi-program experimental evaluation of programs operating in the career pathways framework. The individual PACE evaluations began enrolling study participants between November 2011 and January 2013, depending on the program; the last study participants were enrolled between October 2013 and December 2014. The PACE programs are listed in the text box.

- **Bridge to Employment in the Healthcare Industry (BTH)**, San Diego Workforce Partnership, County of San Diego, CA*
- **Carreras en Salud (CES)**, Instituto del Progreso Latino, Chicago, IL^
- **Health Careers for All (HCA)**, Workforce Development Council of Seattle–King County, Seattle, WA*
- **Integrated Basic Education and Skills Training (I-BEST) program** at three colleges (Bellingham Technical College, Everett Community College, and Whatcom Community College), WA
- **Patient Care Pathway Program (PCPP)**, Madison College, Madison, WI
- **Pathways to Healthcare (PTH)**, Pima Community College, Tucson, AZ*
- **Workforce Training Academy Connect (WTAC)**, Des Moines Area Community College, Des Moines, IA
- **Valley Initiative for Development and Advancement (VIDA)**, Lower Rio Grande Valley, TX
- **Year Up (YU)**, Atlanta, Bay Area, Boston, Chicago, National Capital Region, New York City, Providence, Greater Seattle

*Programs funded through ACF's Health Profession Opportunity Grants (HPOG) Program.

^Program partially HPOG funded.

The PACE project has previously reported on the impact of these programs at 18 months after random assignment and at three years.[1] This work has been guided by an overall evaluation plan (Abt Associates 2014), the analysis plan for the first round of reports (Abt Associates 2015), and the analysis plan for the three-year reports (Judkins, Fein, and Buron 2018). Reports on the impacts of these programs at six years (and for one of the programs at seven years) are published at the ACF website devoted to the Career Pathways Long-Term Outcomes Study.[2] These reports are guided by the analysis plan for the six-year reports (Fein, Judkins, and Buron 2021).[3]

Four Promising Programs

Although all nine PACE programs include at least some components of the career pathways framework, which components included and the intensity of their implementation both vary—as do target populations and focal occupations and industries. Because the programs do not represent a single common programmatic approach, findings were reported separately for each of the nine programs at 18 months and at three years, as well as in cross-site reports at each time point (Gardiner and Juras 2019; Juras and Buron 2021). However, resources for a six-year follow-up survey were only available for a subset of PACE programs. Four of the programs **(Carreras en Salud, I-BEST, VIDA,** and **Year Up)** were deemed to be sufficiently promising to justify the costs of a long-term follow-up survey and separate six-year report.[4] An abbreviated set of six-year findings for the other five programs are included in a cross-site report covering all nine PACE programs as well as the Health Profession Opportunity Grants (HPOG 1.0) Program evaluation.[5]

From a methodological perspective, the evaluations of these programs have much in common, including most items of the survey instrument used for follow-up at six years. The consolidated set of appendices in this volume cover methods used for all nine PACE programs. Due to the

---

[1] The 18-month reports are published at https://www.acf.hhs.gov/opre/project/pathways-advancing-careers-and-education-pace-2007-2018 and the three-year reports are published at https://www.acf.hhs.gov/opre/project/career-pathways-intermediate-outcomes-cpio-study. The six-year reports are published at https://www.acf.hhs.gov/opre/project/career-pathways-long-term-outcomes-study-2016-2021.

[2] For more information on this project, see https://www.acf.hhs.gov/opre/project/career-pathways-long-term-outcomes-study-2016-2021.

[3] https://www.acf.hhs.gov/opre/report/pace-six-year-follow-up-analysis-plan

[4] The decision about which programs were promising enough to warrant follow-up at six years was made in 2017, after drafting of the 18-month reports but before data from the three-year survey were available. Earnings impacts as far out as 14 quarters past the quarter of random assignment were estimated, depending on when study recruitment ended at each program. At the time, only Year Up showed clear evidence of earnings impacts. Carreras en Salud, I-BEST, and VIDA were added to the list of programs for long-term follow-up because of their short-term education impacts, program maturity, and reputation.

[5] The HPOG Program has funded two rounds of grants: HPOG 1.0 in 2010 and HPOG 2.0 in 2015. The six-year cross-site report will consider only programs funded in the first round included in the HPOG 1.0 Impact Study, conducted at about the same time as PACE. The second round of HPOG is being evaluated separately. More information about ACF-funded evaluations of HPOG is available at https://www.acf.hhs.gov/opre/project/health-profession-opportunity-grants-hpog-research-and-evaluation-portfolio.

staggered schedule for publication of the six-year PACE reports, estimates in these appendices may not agree exactly with those in the reports. Survey-based estimates should be stable, but plans to use updated administrative data at VIDA and Year Up mean that many impacts will be (slightly) different than those shown here.[6] Methods used for the evaluation of HPOG 1.0 at six years are reported separately in Litwok, Walton, and Peck (2021).

In This Appendix Volume

The PACE evaluations assessed impacts on a range of outcomes aligned with the career pathways theory of change, including participants' educational progress; credential receipt; career confidence and skills; employment, job quality, and earnings; and general well-being. The **PACE Evaluation Methods** box (pg. viii) briefly describes the methods that the nine studies used to estimate impacts. The appendices in this volume provide much more detail about evaluation methods organized as follows:

A. Baseline Characteristics and Adjustments

B. Six-Year Survey Data

C. National Student Clearinghouse Data

D. Unemployment Insurance Wage Detail

E. Sensitivity Analyses

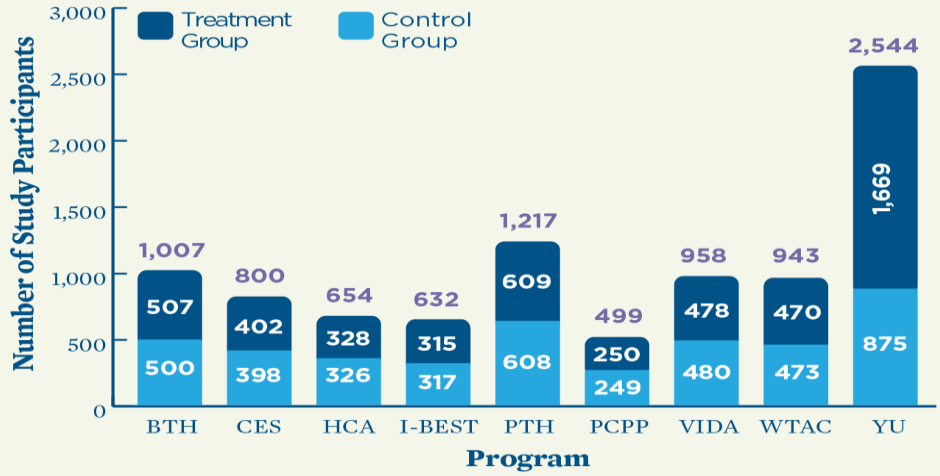
F. Treatment of Outliers

[6] There might be slight differences with estimates for other program as well because sensitivity analyses for all programs were run on NDNH data refreshed in June of 2021, after core tables had been produced for some of the programs based on March 2021 NDNH data.

PACE Evaluation Methods

All nine impact evaluations in the PACE project used experimental research designs to assess impacts of their interventions. For each evaluation, its research team randomly assigned eligible local applicants to either a *treatment group* allowed to access the intervention or a *control group* that could not access the intervention but could access any other trainings, services, and supports available in the community.

The PACE evaluations randomly assigned the following numbers of study participants:



Key: BTH=Bridge to Employment in the Healthcare Industry. CES=Carreras en Salud. HCA=Health Careers for All. I-BEST=Integrated Basic Education and Skills Training. PTH=Pathways to Healthcare. PCPP=Patient Care Pathway Program. VIDA=Valley Initiative for Development and Advancement. WTAC=Workforce Training Academy Connect. YU=Year Up.

The PACE evaluations estimated impacts for each of the nine programs separately. Each evaluation estimated impacts of the intervention as the difference between the treatment group's mean outcomes and the control group's mean outcomes. The control group's experiences represented what would have been absent the intervention.

This is an intent-to-treat design, which estimates the impact of *being offered* access to training and services, as opposed to the impact of *receiving* training and services. Such designs assess whether the treatment group members obtained better outcomes from having access to the intervention than the outcomes they could have obtained without the intervention. Participants in PACE programs chose whether to actually use the services they were offered.

Each program's theory of change identified priority outcomes and time horizons for expected impacts on those outcomes. Each program research team used that program's theory of change to identify one or more *confirmatory outcomes* that best measured the program's effectiveness six (or seven) years after random assignment. All nine evaluations had a confirmatory outcome related to labor market success, and five had an additional confirmatory outcome related to educational attainment.

Additional impact study research questions were intended to generate secondary and exploratory evidence on program effectiveness that could be used to guide future research.

Appendix A: Baseline Characteristics and Adjustments

This appendix starts with a description of the specification for baseline characteristics, including our approach to handling missing values (Section A.1). The next section explains how the analyses control for these covariates in estimating impacts (Section A.2). This material is basically unchanged from the appendices to the three-year reports for PACE such as Judkins, Litwok and Gardiner (2020).

A.1 Details on Baseline Covariates

Exhibit A-1 shows the specifications and data sources for baseline covariates. Item nonresponse rates on these covariates were generally low, less than 5 percent. Exceptions included parental college attendance (10 percent), race (9 percent), Hispanic ethnicity (8 percent), family income (9 percent), receipt of food assistance (7 percent), and receipt of other public assistance (10 percent).

Exhibit A-2 shows the specifications for baseline variables used to define subgroups to study the variation of Year Up impacts in its six-year report.

The team imputed values for missing covariates using SUDAAN®/IMPUTE, a weighted hotdeck imputation procedure (Research Triangle Institute 2012). This process replaced each missing value with an observed response from a similar case. Within specified strata, the process random-matched cases with missing values to cases with reported values then copied over the reported value to the case where the value was missing. The strata represented a cross-classification of treatment-control status, National Student Clearinghouse (NSC)-reported enrollment status (*some* or *none*),[7] NSC-reported credential award (*some* or *none*), and number of months of NSC-reported enrollment.

Exhibit A-1: Operationalization of Baseline Measures Used as Covariates in Regression-Adjusted Impact Estimates

Variable Description	Operationalization Details	Data Source(s) (Survey Instrument: Survey Item Number)
Demographic Background		
Age	Categorical measure: Under 21 21-24 25-34 35+ [a]	BIF: B2_dob RABIT: R_RA_Date_Assigned
Gender	Binary variable: 1 if female 0 if male	BIF: B7

7 NSC has information on monthly enrollment and many credentials for 96 percent of college students. https://nscresearchcenter.org/workingwithourdata/

Variable Description	Operationalization Details	Data Source(s) (Survey Instrument: Survey Item Number)
Race/ethnicity	Categorical measure: Hispanic, any race Black, non-Hispanic White, non-Hispanic[a] Another race, non-Hispanic	BIF: B9
Family structure	Categorical measure: Spouse/partner, with children Spouse/partner, without children Single, with children[a] Single, without children (Only biological and adopted children of randomized participant included here. Stepchildren, grandchildren, younger siblings, and other children not included.)	BIF: B13
Living with own parents	Binary variable: 1 if living with own parent(s) 0 otherwise (Presence of parents of spouse/partner not considered)	BIF: B13
Educational Background		
Parent attended college	Binary variable: 1 if either parent attended college 0 otherwise	BIF: B21
Usual high school grades	Categorical measure: Mostly A's Mostly B's Mostly C's or below[a]	BIF: B23
Highest level of education completed	Categorical measure: No college[a] Less than 1 year of college credit 1+ years of college credit Associate degree or above	BIF: B17
Career Knowledge		
Career Knowledge Index (average of items)	Proportion of responses to seven questions about career orientation and knowledge to which respondent answered, "strongly agree." Missing if four or more of seven responses blank.	SAQ: S13
Psycho-Social Indices		
Academic discipline[b]	Average of 10 items (scale ranging 1=strongly disagree to 6=strongly agree) after reversing responses to negatively phrased items. Missing if 7 or more of 10 responses blank.	SAQ: S11a
Training commitment[c]	Average of 10 items (scale ranging 1=strongly disagree to 6=strongly agree) after reversing responses to negatively phrased items. Missing if 7 or more of 10 responses blank.	SAQ: S11b
Academic confidence[d]	Average of 12 items (scale ranging 1=strongly disagree to 6=strongly agree) after reversing responses to negatively phrased items. Missing if 9 or more of 12 responses blank.	SAQ: S11d
Emotional stability[e]	Average of 12 items (scale ranging 1=strongly disagree to 6=strongly agree) after reversing responses to negatively phrased items. Missing if 9 or more of 12 responses blank.	SAQ: S11e
Social support[f]	Average of 10 items (scale ranging 1=strongly disagree to 4=strongly agree). Missing if 9 or more of 10 responses blank.	SAQ: S12

Variable Description	Operationalization Details	Data Source(s) (Survey Instrument: Survey Item Number)
Resource Constraints (Financial)		
Family income in past 12 months	Categorical measure: Less than $15,000 $15,000-$29,999 $30,000+ [a]	BIF: B27
Received food assistance (WIC/SNAP) in past 12 months	Binary variable: 1 if yes 0 if no	BIF: B26b
Received public assistance or welfare in past 12 months	Binary variable: 1 if yes 0 if no	BIF: B26c
Financial hardship in past 12 months	Binary variable: 1 if yes to ever missed rent/mortgage payment in prior 12 months or reported generally not having enough money left at the end of the month to make ends meet over the last 12 months. 0 if otherwise	SAQ: S8, S9
Resource Constraints (Time)		
Current work hours	Categorical measure: 0-19[a] 20-34 35+	BIF: B24
Expected work hours in next few months	Categorical measure: 0-19[a] 20-34 35+	SAQ: S2
Expecting to attend school part-time if accepted	Binary variable: 1 if yes, 0 if no	SAQ: S1
Life Challenges		
Frequency of situations interfering with school, work, job search, or family responsibilities	Average of six items of frequency of problems in past 12 months (childcare, transportation, alcohol or drug use, health, family arguments, physical threats). Scale ranges from 1=never to 5=very often. Missing if 4 or more of 6 responses blank.	SAQ: S15
Stress[f]	Average of four items about feeling in control of important things and able to handle personal problems (scale 1=never to 5=very often over the past month) after reversing responses to negatively phrased items. Missing if 3 or more of 4 responses blank.	SAQ: S14

Key: BIF = Basic Information Form. RABIT = Random Assignment and Baseline Information Tool. SAQ = Self-Administered Questionnaire. SNAP = Supplemental Nutrition Assistance Program. WIC = Special Supplemental Nutrition Program for Women, Infants, and Children.

[a] Category omitted in creating binary (dummy) variables for regression-adjustment models.

[b] Modified version of the Academic Discipline scale in the Student Readiness Index (SRI), a proprietary product of ACT, Inc.; Le et al. (2005). Further validation in Peterson, Casillas, and Robbins (2006).

[c] Modified version of Commitment to College scale in the Student Readiness Index (SRI), a proprietary product of ACT, Inc.; Le et al. (2005). Further validation in Peterson, Casillas, and Robbins (2006).

[d] Modified version of the Academic Self-Confidence scale in the Student Readiness Index (SRI), a proprietary product of ACT, Inc.; Le et al. (2005). Further validation in Peterson, Casillas, and Robbins (2006).

[e] Modified version of the Emotional Control scale in the Student Readiness Index (SRI), a proprietary product of ACT, Inc.; Le et al. (2005). Further validation in Peterson, Casillas, and Robbins (2006).

[f] Modified version of the Social Provisions Scale; Cutrona and Russell (1987). Original scale has 24 items. This short version developed by Hoven (2012).

[g] Cohen, Kamarck, and Mermelstein (1983).

Exhibit A-2: Operationalization of Subgroups Studied in the Year Up Six-Year Report

Variable Description	Operationalization Details	Data Source(s) (Survey Instrument: Survey Item Number)
Demographic Background		
Age	Under 20 20-22 23-24	BIF: B2_dob RABIT: R_RA_Date_Assigned
Gender	Male Female	BIF: B7
Race/ethnicity	Hispanic, any race Black, non-Hispanic White or other non-Black race, non-Hispanic	BIF: B9
Educational Background		
Usual high school grades	Mostly A's and B's Mostly C's or below	BIF: B23
Highest level of education completed	High school Some but less than 1 year of college 1+ years of college	BIF: B17
Psycho-Social Indices		
Tertiles of training commitment[a]	After forming scale as defined in Exhibit A-1, split as closely as possible to equal thirds	SAQ: S11b
Tertiles of depression[b]	First formed average of 9 items (scale ranging 1=rarely felt symptom to 4=felt symptom most of the time). Missing if 6 or more of 9 responses blank. Then split as closely as possible to equal thirds.	SAQ S16
Resource Constraints (Time)		
Expected work hours in next few months	0-9 10-29 30+	SAQ: S2
Life Challenges		
Frequency of situations interfering with school, work, job search, or family responsibilities	Average of 6 items of frequency of problems in past 12 months (childcare, transportation, alcohol or drug use, health, family arguments, physical threats). Scale ranges from 1=never to 5=very often. Missing if 4 or more of 6 responses blank.	SAQ: S15

Key: BIF = Basic Information Form. RABIT = Random Assignment and Baseline Information Tool. SAQ = Self-Administered Questionnaire. SNAP = Supplemental Nutrition Assistance Program. WIC = Special Supplemental Nutrition Program for Women, Infants, and Children.

[a] Modified version of Commitment to College scale in the Student Readiness Index (SRI), a proprietary product of ACT, Inc.; Le et al. (2005). Further validation in Peterson, Casillas, and Robbins (2006).

[b] Short-form version of the Center for Epidemiologic Studies Depression (CES-D) scale developed by Santor and Coyne (1997).

A.2 Regression Adjustment

This section describes the regression adjustment approach used to improve precision and minimize effects of sampling error on impact point estimates. In a rigorous evaluation, random assignment ensures that if the sample size is large enough, differences in average potential outcomes between the treatment and control groups will become vanishingly small so that any observed differences in average outcomes across the two groups must almost certainly be the result of treatment.[8] Even when sample sizes are modest (as is true here), random assignment ensures that differences in average potential outcomes between the treatment and control groups arise from chance rather than biases of program operators or program evaluators. This means that unbiased estimates of the effects of treatment can be obtained by simply comparing average outcomes across the treatment and control groups. Moreover, it is easy to run formal tests of the hypothesis that the program has no effect (and that therefore the observed difference in mean outcomes is the result of those accidental imbalances in potential outcomes across the two groups).

Despite these favorable properties of analysis based on simple comparisons of observed means, use of regression adjustment can reduce the impact of accidental imbalances in potential outcomes across the groups, thereby increasing power to detect small program impacts (Lin 2013). To achieve this benefit, the variables used in the regression adjustment must be predictive of potential outcomes. Including other variables will increase the variance on the estimated program impact rather than decreasing it.

Opinions and practice differ on how strong the evidence for correlation between a baseline variable and the outcome must be before it makes sense to include the baseline variable in the regression adjustment.[9] Some favor a lean approach, including just those baseline variables that have a demonstrated strong relationship to the outcome; others favor a more comprehensive approach, including all baseline variables that have a plausible theoretical relationship to outcomes of interest, believing that doing so generally bolsters confidence in study findings (Tukey 1991).

Given demands to minimize burden on participants, all measured PACE baseline variables have at least plausible relationships to PACE outcomes measured at six months, but some baseline variables have been discovered to have only weak empirical relationships with PACE outcomes. Moreover, one could combine the directly measured characteristics into a limited number of interactions. So some judgment must be exercised about which covariates to include in regression adjustments and which to exclude.

Opinions and practice also differ on how much to customize decisions about covariate inclusion across outcomes in evaluations (like these nine PACE evaluations) with multiple outcomes. A single uniform set of decisions promotes transparency, making it easier for readers to understand the procedure, whereas a more customized approach is likely to improve variances for at least some outcomes given that the correlation between a covariate and an outcome will vary by outcome.

---

8 *Potential outcomes* are a central concept in the Neyman-Rubin causal model (Holland 1986). In the model, each person has an innate pair of possible outcomes: one if treated and the other if not treated. Only one of the two potential outcomes is ever observed for each person. The average difference in potential outcomes across a specific population is said to be the local average treatment effect (LATE), or more simply just the effect of treatment, with the context making clear the population to which it applies and supplemental analyses exploring whether the effect is homogenous within that population.

9 For a current review of practice, see Ciolino et al. (2019).

In preliminary analyses for the first round of PACE reports at 18 months, the team used a fairly comprehensive approach with a uniform set of decisions but discovered that this approach was causing the variances on adjusted impacts to be larger than the variances on unadjusted impacts. The discovery prompted a switch to a different approach for the first round of reports, which ultimately proved not to work as well as hoped (Judkins 2019). In response, the team developed a new approach for the intermediate (three-year) round of PACE reports.[10] The long-term (six-year) reports also use that new approach.

The new approach emphasizes transparency and control on imbalanced covariates, while still trying to maximize precision as far as possible given those priorities. Details follow.

Equation (A.1) below shows the conventional regression-adjustment model:

$$Y_i = X_i\beta + \delta T_i + e_i \tag{A.1}$$

where Y_i is the outcome; X_i is a row vector of baseline characteristics (hereafter referred to as covariates); β is the vector of parameters indicating the influence of each covariate on the outcome; δ is the effect of treatment; T_i is a 0/1 dummy variable indicating treatment group membership; and e_i is an error term. We fit models of this sort using SAS®/SURVEYREG, a procedure that uses a robust estimator of the variance of $\hat{\delta}$ and can accommodate the required nonresponse-adjustment weights for survey-measured outcomes. (See Appendix Section B.4 for a discussion of nonresponse-adjustment weights.)

This method is known as ordinary least squares (OLS) and has excellent properties when the sample size is many times larger than the number of baseline characteristics used as covariates (Lin 2013), even when the outcomes are not normally distributed (Judkins and Porter 2016). Estimates of the treatment effect are "asymptotically unbiased." Furthermore, except under conditions discussed in the next few paragraphs, the variance of the estimated treatment effect declines from the simple treatment/control difference-in-mean-outcomes estimator of impact in proportion to the amount of outcome variation explained by the covariates.

Specifically, the relationship between the variance of the estimated treatment effect from the OLS estimation of Equation (A.1) and the explanatory power of the covariates is $var(\hat{\delta}) \approx (1 - R^2)var(\bar{y}_t - \bar{y}_c)$, where R^2 is the proportion of the variance in Y_i explained by the baseline characteristics (X_i) in OLS estimation of Equation (A.2) below:

$$Y_i = X_i\beta + e_i \tag{A.2}$$

However, as mentioned above, when there are a large number of potential covariates, not all of which are useful in predicting every outcome of interest, the effect of adjustment can be the

[10] We considered simply using the same covariates as were used in the prior round of reports but decided to optimize the covariate selections for the current round of reports. Reselecting covariates for this round introduced small changes in impact estimates for three-year outcomes included in the six-year reports, occasionally moving tests slightly above/below significance thresholds and slightly affecting comparability across reports. We assumed the advantage of calibrating to new focal outcomes was worth this cost.

opposite of the intended effect: variances are increased rather than decreased.[11] To avoid unnecessary variance inflation, the analyst needs to drop or otherwise reduce the influence of extraneous covariates that do not have a strong influence on the outcome of interest.

Simulation research (Judkins 2019) showed that dropping (with “backward selection”) or downweighting covariates[12] based on simple analyses of the same data used in the evaluation yields slightly biased estimates of the variance of the estimated treatment effects (but still unbiased estimates of the treatment effect itself).[13] This bias is negative, meaning that the variance estimates are slightly too small, making confidence intervals for impact estimates misleadingly narrow and hypothesis tests too likely to conclude that a positive impact has occurred when the true impact is zero or negative.

To select covariates in a manner that does not compromise variance estimation, we use the relatively recently developed technique “least absolute shrinkage and selection operator” (LASSO) with “10-fold cross-validation.”[14] With the LASSO, the sum of absolute values of the estimated regression coefficients in Equation (A.2) is constrained to be less than a preselected value (the “constraint”). If the value for this constraint is small enough, many coefficients in Equation (A.2) will be forced to zero in order to fit within the cap on the sum of absolute coefficient values and thus can be removed from the list of baseline covariates. The 10-fold cross-validation is used to optimize the value of the constraint, rather than just relying on an arbitrary choice for it.

Details of the procedure are as follows:

(1) With 10-fold cross-validation, the sample (both treatment and control group members) is divided into 10 equal and mutually exclusive random subsamples.

(2) For each of a range of candidate values of the constraint, the LASSO procedure is run to select covariates on a sample in which one of the 10 subsamples has been dropped.

---

11 Mathematically, the presence of extraneous variables causes the coefficients of the true determinants of the outcome to be less accurately estimated. For example, if the best prediction model is $Y = 2X$ but the model is fit with many extraneous covariates, the fit prediction formula could easily end up having coefficients of 1.9 or 2.1 for X instead of the best value of 2. If the wrong slope is used to correct for a treatment-control imbalance in X, the adjusted estimate of impact can be worse than an unadjusted estimate of impact.

12 An example of a method that downweights covariates is the “modified Koch method” developed for and used in the first round of PACE reports (see Judkins, Fein, and Buron 2018; Koch et al. 1998).

13 If the sample size is very large, the estimated variance of the estimated effect of treatment will be nearly unbiased even if the evaluation data are used to cull or downweight extraneous covariates. However, simulations clearly show that PACE sample sizes are not large enough to avoid biased variance estimates if “backward selection” on local data is used to prune covariates or if the modified Koch method is used to downweight extraneous covariates. Accordingly, impact analyses at three and six years for PACE programs are not using the modified Koch method used in the first, short-term round of reports covering the first 18 months of follow-up.

14 See Bühlmann and van de Geer (2011) for a full explanation of these techniques.

(3) The model in Equation (A.2) is fit on the same sample using just the variables selected in the second step for each of the candidate values of the constraint.

(4) The model is used to create out-of-sample predictions of the outcome for everyone in the dropped piece of the sample, and the prediction error $\hat{Y}_i - Y_i$ is measured for each of the candidate values of the constraint.

(5) Steps 2 through 4 are repeated 10 times for each candidate value of the constraint. On each iteration, a different one of the 10 subsamples is dropped. In this manner, out-of-sample prediction errors are obtained for the entire sample.

(6) Mean squared prediction errors across all 10 replicates are then calculated for each of the candidate values of the constraint.

(7) The value of the constraint that minimizes this cross-validated mean squared prediction error and thus captures most of the variation reduction possible with the available covariates is selected as the optimal constraint.[15] Whichever variables have nonzero coefficients in the model for that optimal constraint are used as covariates in the impact regressions. All other baseline characteristics are discarded. All of this is done automatically in SAS®/GLMSELECT. Simulations under PACE-like conditions (in terms of sample sizes and the numbers of covariates) when developing the analysis plan for the entire suite of PACE three-year reports (Judkins, Fein, and Buron 2018) demonstrate that this technique reduces the true variances without biasing variance estimates.[16]

In principle, we could repeat the LASSO with 10-fold cross-validation independently for every outcome for each of the nine PACE programs. But such an approach would produce a different final covariate list for each outcome and program, leading to some loss in transparency and making it harder for outside researchers to replicate the PACE results. At the other extreme, we could run the LASSO just once for each program for the most important confirmatory outcome and then use the resulting set of selected covariates for all impact estimates for the program.

As a compromise between these extremes, we ran the LASSO separately for each of three domains at the four PACE programs with six-year follow-up surveys. At the other five PACE programs, we ran the LASSO separately for only two of the three domains:

- Analyses of **employment and earnings** outcomes that are based on the National Directory of New Hires (NDNH)—at all nine PACE programs.
- Analyses of **educational progress** outcomes (whether based on the three-year survey or NSC records)—at all nine PACE programs.
- Analyses of all **other survey outcomes**, most of which concern personal and family well-being and economic independence, but also include some six-year survey-based

15 One could simply use the LASSO to select covariates with a pre-specified value of the constraint, but the 10-fold cross-validation provides a principled method for selecting the constraint.

16 See Judkins (2019) for additional detail.

measurements of employment and earnings)—only at the four promising PACE programs surveyed at six years.

The pool of potential covariates was the same for all three domains—with one important exception: indicators of pre-baseline earnings based on NDNH data are allowed only in analyses of NDNH-based outcomes.[17]

To identify covariates for this report, we ran the LASSO procedure for the most salient outcome within each measured domain at each of the nine PACE programs.[18] For NDNH analyses, the confirmatory outcome is *average quarterly earnings for the 23rd and 24th quarters after randomization* (Q23, Q24), so that was a natural choice for the outcome around which to optimize covariate selection. In the educational progress domain, the most salient outcome for all sites (except Pathways to Healthcare) was *receipt of a college credential after eight or more months of FTE college enrollment by Q24*.[19] For the third domain (all others), which we analyzed only at the four surveyed sites, we selected *household income in the month prior to survey interview* as the target outcome for the LASSO.[20] We made these choices prior to reviewing any impact estimates.

In addition to covariates based on the above procedures, regression models included covariates for which baseline distributions differ for treatment and control group members at the 5 percent level.[21] These are referred to in the balance of this appendix as being "out of balance."

Exhibit A-3 displays which baseline measures were used as covariates in regression models by domain and also by site. A dollar sign ($) indicates that it was used for all NDNH outcomes. The icon (graduation cap) indicates that the baseline measure was used in the educational

17 This is because we analyzed survey outcomes on Abt's secure server rather than on the ACF secure server. Though both systems have very high security procedures, agreements with the Office of Child Support Enforcement (OCSE) permit the NDNH data to reside only on the ACF secure server. It would have been possible to analyze all survey outcomes on the ACF secure server, but doing so would have significantly burdened the study's analytic operations with little benefit. It would also prevent us from analyzing survey data for people whose names and Social Security numbers do not properly match according to OCSE.

18 Selection started with the set of baseline covariates used in the analyses of follow-up data at 18 months after random assignment for the first round of PACE reports (shown in Exhibit A-3).

19 For Pathways to Healthcare, prior analysis rounds used records from its host college (Pima Community College) to measure *receipt of a college credential requiring at least one year of study to earn*, with NSC-based imputation for attendance at other colleges. Over time, attendance at other colleges has increased, rendering it untenable to rely on imputation for an important confirmatory outcome. Meanwhile, the research team learned that dates on Pima records provided to NSC do not support accurate measures of periods of enrollment (Judkins, Litwok, and Gardiner 2020, Section D.3). It is thus not possible to measure enrollment-duration-dependent credential receipt well for this site. Therefore, *receipt of a college degree by Q24* is the target outcome for the LASSO in the educational progress domain for Pathways to Healthcare.

20 As discussed in Appendix Section B.3, we utilized multiple imputation for several outcomes, including household income. For the purposes of covariate selection, we used the first imputation as the target variable for the LASSO.

21 Baseline balance was assessed prior to imputation of missing data.

progress domain (for both NSC and survey outcomes in this domain). The icon ☤ (ear of corn) is meant to suggest family well-being and indicates that the covariate was used for all other survey domains. An “O” indicates that the measure was out of balance at baseline at the site and therefore used as a regression covariate for outcomes in all domains at that site.

Exhibit A-3: Selected Baseline Measures for Regression Adjustment, by PACE Site

Baseline Measure	PACE Site								
	BTH	CES	HCA	I-BEST	PCPP	PTH	WTAC	VIDA	YU
Age under 20									🎓
Age 20 to 24	🎓	🎓 🌷				🎓	🎓	🌷	
Age 25 to 34	$			🎓	$				
At least one parent with some college		🌷		🎓	🎓				
Mostly A grades in high school	🎓	🎓		🎓		🎓		$ 🎓	🎓
Mostly B grades in high school				🎓 🌷					🎓
Some college but less than 1 year			O				🎓	O	$
1+ years of college	🎓	$ 🎓 🌷	🎓 O	🎓		🎓	🎓	$ 🎓 🌷 O	$ 🎓 🌷
Associate degree or higher	$ 🎓	🎓	O			$ 🎓		$ 🎓 🌷 O	$ 🎓
Female	$	🌷	$ 🎓 O	$ 🎓 🌷				$ 🌷	🎓 🌷
Hispanic, any race				🎓	O				
Black, non-Hispanic				🎓	🎓 O		🎓		$ 🌷
Another race, non-Hispanic	$				O				$ 🎓
Not living with spouse/partner or children						$	O	🎓 🌷 O	
Not living with spouse/partner, living with children				🌷			O	🎓 🌷 O	
Living with spouse/partner and children							O	O	
Living with parents	🎓					$			🎓
Family income last year below $15,000	🌷	🌷			🎓 O				$ 🌷
Family income last year $15,000 to $29,999					O				🌷
Received food assistance in past 12 months	$	🎓		$ 🎓 🌷	🎓	$		🌷	$
Received public assistance or welfare in past 12 months			🎓	🌷				🌷	🎓
Financial hardship in past 12 months	🎓			🎓				🎓	🎓
Current work hours 20 to 34 per week	🎓 $			$		$		🎓 🌷 O	🌷

Baseline Measure	PACE Site								
	BTH	CES	HCA	I-BEST	PCPP	PTH	WTAC	VIDA	YU
Current work hours 35+ per week	$	i					$	O	🌱 $
Expected work hours in next few months 20 to 34 per week	$			🎓 🌱	$				🎓
Expected work hours in next few months 35+ per week	🎓 $	🎓	🎓	$		🎓		🎓	
Plan to attend school only part-time if admitted to program									
Academic discipline index	O			🌱			O	$	
Training commitment index									
Emotional stability index				🌱			🎓 O		
Academic self-confidence index		🎓 🌱 O	🎓					🎓 $	
Stress index		🌱							
Live challenges index	O			$ 🌱		$	$	🌱	$
Career knowledge index	O							🌱	🎓
Earnings in 1st quarter prior to randomization		$		$	$		$		$
Earnings in 2nd quarter prior to randomization	$		$	$	$	$	$		
Earnings in 3rd quarter prior to randomization			$					$	
Earnings in 4th quarter prior to randomization							$	$	
Employed in 1st quarter prior to randomization						$		$	
Employed in 2nd quarter prior to randomization									
Employed in 3rd quarter prior to randomization									
Employed in 4th quarter prior to randomization									

Key: $ = Predictive of Q23/24 earnings. 🎓 = Predictive of educational progress. 🌱 = Predictive of household well-being. O = Out of balance at baseline.

Sites: BTH=Bridge to Employment in the Healthcare Industry. CES=Carreras en Salud. HCA=Health Careers for All. I-BEST=Integrated Basic Education and Skills Training. PTH=Pathways to Healthcare. PCPP=Patient Care Pathway Program. VIDA=Valley Initiative for Development and Advancement. WTAC=Workforce Training Academy Connect. YU=Year Up.

Note: All measures defined in Section A.1. Some rows correspond to individual values for categorical variables with three or more levels. This is indicated by shading. Reference categories are omitted.

Appendix B: Six-Year Survey Data

This appendix documents key technical details underlying analyses of the six-year follow-up survey data.[22] Section B.1 documents coding for scales based on follow-up survey data—excluding coding for credential classification, which is documented in Section B.2. Section B.3 describes the imputation process for some missing survey data elements in the construction of outcomes. Section B.4 analyzes survey nonresponse and documents the process we used to build the nonresponse weights used in the impact analysis.

The six-year follow-up survey sought to collect a complete history of new credentials earned since the date of the three-year survey interview (or since the date of study enrollment, for those not interviewed at three years); current employment and conditions of employment; adult well-being (including stress, life challenges, and physical health); economic well-being (including income, debt, financial reserves, health insurance, food security and housing); family formation; and family life.

B.1 Measures Based on Follow-Up Survey Data Except the Earning of Credentials

Exhibits in this section detail the operationalization of survey-based outcomes from closed-ended questions used in impact analyses in the main report. These exhibits also reference the underlying survey questions. Exhibit B-1 provides details on outcomes in the employment and earnings domain. Exhibit B-2 provides similar details on outcomes in the educational progress domain. Exhibits B-3, B-4, and B-5 do the same for the “other” domain comprising economic well-being, family formation, and adult well-being, respectively. Finally, Exhibit B-6 provides details on parenting and child development outcomes.

[22] The full instrument is available at https://www.reginfo.gov/public/do/PRAViewIC?ref_nbr=201802-0970-010&icID=227184.

Exhibit B-1: Details on Specifications for Survey-Based Outcomes in the Employment and Earnings Domain, Six-Year Follow-Up Survey

Outcome	Details on Derivation of Outcome	Six-Year Follow-Up Survey Question(s)
Secondary Outcomes		
Working full-time (35+ hours/week)	Respondent is employed and typically works 35 or more hours per week.	B3, B13, B20
Working in a program target occupation (I-BEST and Year Up only)	Certain SOC codes were deemed to be “targeted” by the I-BEST and Year Up programs; participants who were employed in occupations covered by these SOC codes were considered to be employed in an occupation targeted by I-BEST or Year Up training.[a]	B3, B13, C2, C3, C4, C5
Working in the healthcare field (Carreras only)	Participants who were employed in occupations with SOC codes starting with “29” or “31” were considered to be employed in the healthcare field.	B3, B13, C2, C3, C4, C5
Working in a job offering a full set of benefits	Currently employed in a job offering health insurance, paid vacation, paid holidays, paid sick days, and retirement or pension benefits.	B3, B13, C8
Access to career network	This was a new scale created for PACE at the 18-month follow-up. The 6-item scale counts the number of types of career-supportive relationships in workforce and education settings the respondent can claim. The motivation for creating this scale was the theory that richer social networks are one of the benefits of higher education (e.g., Goldrick-Rab and Sorensen 2010). The outcome counts the number of “Yes” responses to each item, and thus ranges from 0 to 6. Missing if 4 or more of 6 responses are blank. *Say you need advice or help in taking a next step on a career pathway of interest to you. Please tell me if there is anyone you’d be comfortable turning to:* • Who has a college degree? • Who is currently going to college? • Who works at a local college, either as a teacher or staff member providing help to applicants or students? • Who works for a local community organization helping people find education and training, work, and related supports? • Who works in an occupation of interest to you? • Who has a management job in a work setting matching your career interests?	E1
Exploratory Outcomes		
Hours working per week	Respondents were asked how many hours, on average, they are working per week in their current job. Responses to this question allowed the calculation of the percentage of respondents not currently employed, working 1-19 hours per week, 20-34 hours per week, and 35 or more hours per week, along with average/median weekly hours among survey respondents.	B3, B13, B20
Hourly wages (if employed)	Respondents who reported employment at the time of the survey were asked their typical pre-tax hourly wage. Responses to this question allowed the calculation of the percentage of employed respondents earning $1-$9, $10-$14, $15-$19, $20-$29, $30-$39, and $40+ an hour, as well as median and average hourly wages.	B3, B13, B18, B18a, B19, B19a, B20
Weekly earnings	Continuous measure of weekly earnings defined by multiplying the hourly wage by the number of hours typically worked per week in current job. Median and average weekly earnings reported.	B3, B13, B18, B18a, B19, B19a, B20

Outcome	Details on Derivation of Outcome	Six-Year Follow-Up Survey Question(s)
Working in a job at or above $15/hour	Currently employed in a job with an hourly wage above $15 per hour.	B3, B13, B18, B18a, B19, B19a, B20
Working in a job at or above $20/hour	Currently employed in a job with an hourly wage above $20 per hour.	B3, B13, B18, B18a, B19, B19a, B20
Working in a job at or above $25/hour	Currently employed in a job with an hourly wage above $25 per hour.	B3, B13, B18, B18a, B19, B19a, B20
Working in a job offering health insurance	Currently employed in a job offering health insurance.	B3, B13, C8_1
Working in a job offering paid vacation	Currently employed in a job offering paid vacation.	B3, B13, C8_2
Working in a job offering paid holidays	Currently employed in a job offering paid holidays.	B3, B13, C8_3
Working in a job offering paid sick days	Currently employed in a job offering paid sick days.	B3, B13, C8_4
Working in a job offering retirement or pension benefits	Currently employed in a job offering retirement or pension benefits.	B3, B13, C8_5
Perceived career progress	This was a new scale created for PACE at the 18-month follow-up. It is a 3-item scale of self-assessed career progress. It was designed specifically to measure a respondent's sense of progress in a career pathways program as described by Fein (2012). Response categories range from 1=strongly disagree to 4=strongly agree. • I am making progress toward my long-range educational goals • I am making progress toward my long-range employment goals • I see myself on a career path	D10
Received any promotions in the last 3 years	Respondent has received any promotions or changed employers to obtain a promotion in the last 3 years.	B3, B13, B21
Changed employers for better job in last 3 years	Respondent has changed employers to take a new job at a higher level or to take a new job with a higher pay rate in the past 3 years.	B3, B13, B21, B23d, B23e
Career connected	Fully engaged in career-related employment or education: either employed full-time, training full-time, or both employed and training at least part-time.	B13, B20, D2, D3
Occupational sector (if employed)	Respondents who reported employment at the time of the survey were asked to describe the typical activities and duties involved in their current job. This information was converted into an occupational sector classification.	B3, B13, C2, C3, C4, C5

Key: SOC = U.S. Department of Labor Standard Occupational Classification.

[a] The SOC codes targeted by the Year Up program are those starting with 11, 13, 15, 41-3, 41-4, 43-3, 17-2, 27-3, 27-4, and six-digit codes 41-1011, 41-9099, 43-4051, 43-9011, 43-9111, 49-2011, and 51-9061. The SOC six-digit codes targeted by the I-BEST program are 49-3023, 47-2111, 31-1131, 51-4041, 51-4120, 49-2090, and 43-9060.

Exhibit B-2: Details on Specifications for Survey-Based Outcomes in the Educational Progress Domain Other than Earning of Credentials, Six-Year Follow-Up Survey

Outcome	Details on Derivation of Outcome	Six-Year Follow-Up Survey Question(s)
Other Exploratory Outcomes		
Ever enrolled in education/training in follow-up Years 4-6	Respondent was enrolled in any education or training at any time during the last 3 years. Imputed for those who did not respond to 3-year survey.	D2, D2a
Enrolled in education/training at a college at time of survey	Enrolled at a community or technical college or 4-year college/university at the time of the survey.	D2, D3a
Enrolled in education/training at another education/training institution at time of survey	Enrolled at a school that is not a college, a community organization, an employer, or other at the time of the survey.	D2, D3a
Enrolled in education/training at a school that is not a college at time of survey	Enrolled at an adult education/adult high school/community school or a private school/company that provides training at the time of the survey.	D2, D3a
Enrolled in education/training at a community organization at time of survey	Enrolled at a community based/nonprofit organization, state unemployment/employment office, or One-Stop career center at the time of the survey.	D2, D3a
Enrolled in education/training at an employer at time of survey	Enrolled at place of employment at the time of the survey.	D2, D3a
Enrolled in education/training at other place at time of survey	Enrolled in online classes or someplace else at the time of the survey.	D2, D3a
Enrolled in any education/training program at time of survey	Enrolled in education or training at the time of the survey.	D2
Enrolled full-time at time of survey	Enrolled in education or training at least 12 hours per week at the time of the survey.	D2, D3
Enrolled part-time at time of survey	Enrolled in education or training less than 12 hours per week at the time of the survey.	D2, D3
Highest education level was bachelor's or higher	Highest level of education completed was a bachelor's degree or higher at the time of the survey.	D1
Highest education level was associate degree	Highest level of education completed was an associate degree at the time of the survey.	D1
Highest education level was 1+ years of college, no degree	Highest level of education completed was 1 or more years of college but no degree at the time of the survey.	D1
Highest education level was some college, but less than 1 year	Highest level of education completed was some college, but less than 1 year at the time of the survey.	D1
Highest education level was high school diploma or equivalent	Highest level of education completed was a high school diploma or equivalent at the time of the survey.	D1

Exhibit B-3: Details on Specifications for Survey-Based Outcomes in the Economic Well-Being Domain, Six-Year Follow-Up Survey

Outcome	Details on Derivation of Outcome	Six-Year Follow-Up Survey Question(s)
Secondary Outcomes		
Able to handle a financial emergency of $400	Respondent is able to handle a financial emergency of $400 or more using cash or money in their checking/savings account.	G21
Total debt	Total debt from student loans in own or parent's name and other "unsecured" debt (e.g., credit cards). Excludes "secured" debts (e.g., mortgages and car loans). Debts in the name of spouse or partner were included.	G11, G11a, G15, G15a, G16
Financial distress	This scale is an expanded version of the financial hardship measure used in the18-month follow-up survey. It measures the extent of financial distress by counting the number of domains (out of a total of 9) for which respondent reported signs of financial distress. The categories are troubles paying bills (rent/mortgage, gas/oil/electricity), utility disconnects (gas/electric/oil, telephone), delayed healthcare, delayed dental care, delayed prescription drug procurement, not having enough to eat (sometimes or often), and not having enough money to make ends meet at the end of the month.	G17, G18, G19
Receives means-tested public benefits	Respondent or someone else in their household received TANF, SNAP, WIC, Medicaid, subsidized childcare, Section 8 or Public Housing, LIHEAP, or FRPL in the month prior to interview.	G1a, G1b, G1c, G1e, G1f, G1g, G1h, G1i
Exploratory Outcomes		
Household income	Respondents were first asked to provide an open-ended amount for the prior month, specifically excluding income tax refunds, where the "household" was clarified to include anyone who lived in the household for at least half of the prior month. If no answer was given, the respondent was asked to choose one of 7 bracketed amounts. Item nonresponse was multiply imputed. Exact amounts were also multiply imputed for people who chose a bracket. People who lived alone were not asked this question; instead, their personal income was assumed to equal the household income.	G8, G8a
Personal income	Respondents were first asked to provide an open-ended amount for the prior month, specifically excluding income tax refunds. If no answer was given, the respondent was asked to choose one of 7 bracketed amounts. Item nonresponse was multiply imputed. Exact amounts were also multiply imputed for people who chose a bracket.	G7, G7a
Able to handle a financial emergency of $200	Respondent is able to handle a financial emergency of $200 using cash or money in their checking/savings account.	G21
Student debt	Students were asked about personal borrowing to go to school since randomization. For those who had difficulty answering the question about the exact amount, a categorical response option was offered. Item nonresponse was multiply imputed for Year Up. Exact amounts were also multiply imputed for people who chose a bracket. For other sites, first imputation used due to very low rates of missing data.	G11, G11a

Outcome	Details on Derivation of Outcome	Six-Year Follow-Up Survey Question(s)
Parental student debt	Respondents were asked about borrowing by parents on behalf of the student to go to school since randomization. For those who had difficulty answering the question about the exact amount, a categorical response option was offered. Item nonresponse was multiply imputed for Year Up. Exact amounts were also multiply imputed for people who chose a bracket. For other sites, the first imputation was used due to very low rates of missing data.	G15, G15a
Other debt	Respondents were asked about debt other than student debt, title debt (car loan), and mortgage debt. Exact amounts were imputed from categorical responses.	G16
Has health insurance coverage	Respondent currently has health insurance, either through employer or another source.	G22, G23
Sometimes/often not enough to eat	Respondent sometimes or often does not have enough to eat.	G18
Neither owns/rents home or apartment	Respondent does not rent or own their home or apartment.	G24
Lived with either a friend or relative sometime during last six months for lack of income	Respondent lived with a friend or relative during the last 6 months because they could not find or afford a place of their own.	G25
Receipt of Unemployment Insurance or Workers Compensation	Respondent or someone else in their household received Unemployment Insurance or Workers Compensation in the month prior to interview.	G1d
EITC receipt	Respondent claimed the Earned Income Tax Credit for the year prior to the interview.	G9
Receipt of money from family or friends	Respondent or someone else in their household received money from family or friends who did not live with them at least half of the time in the month prior to interview.	G1k

Key: EITC = Earned Income Tax Credit. FRPL = free or reduced-price lunch. LIHEAP = Low Income Home Energy Assistance Program. SNAP = Supplemental Nutrition Assistance Program. TANF = Temporary Assistance for Needy Families. WIC = Special Supplemental Nutrition Program for Women, Infants, and Children.

Exhibit B-4: Details on Specifications for Survey-Based Outcomes in the Family Formation Domain, Six-Year Follow-Up Survey

Outcome	Details on Derivation of Outcome	Six-Year Follow-Up Survey Question(s)
Exploratory Outcomes		
Family formation	Respondents were asked who currently lives in their household (i.e., whether they live with their parents, spouse/partner, children, etc.).	F1
Birth since random assignment/Current pregnancy	Respondents were asked whether they or their partner has had a child since random assignment or whether they or their partner is currently pregnant.	F6, F7

Exhibit B-5: Details on Specifications for Survey-Based Outcomes in the Adult Well-Being Domain, Six-Year Follow-Up Survey

Outcome	Details on Derivation of Outcome	Six-Year Follow-Up Survey Question(s)
Exploratory Outcomes		
Life Challenges Index	A new scale adapted for PACE from a longer instrument by Kessler et al. (1998). Average of 4 items of frequency of situations that interfered with school, work, job search, or family responsibilities. The response categories ranged from 1=never to 5=very often. Missing if 3 or more of 4 responses are blank. • Childcare arrangements • Transportation • Alcohol or drug use • An illness or health condition	E3
Stress Index	Existing scale from Cohen, Kamarck, and Mermelstein (1983). This scale was first used in the PACE Basic Information Form and has since been included in the follow-up instruments. The 4-item scale captured perceived stress. The response categories ranged from 1=never to 4=very often. Missing if 3 or more of 4 responses are blank.	E4
Social Support Index	Existing scale from Hoven (2012). The 10-item scale response categories ranged from 1=strongly disagree to 4=strongly agree. It is a short-form version of the Social Provisions Scale (Cutrona and Russell, 1987), a scale that has 24 items. Missing if 7 or more of 10 items are blank.	E2
Physical Health	Respondents reporting excellent, very good, fair, or poor health.	E5

Exhibit B-6: Details on Specifications for Survey-Based Outcomes in the Parenting and Child Development Domain, Six-Year Follow-Up Survey (Carreras en Salud and VIDA only)

Outcome	Details on Derivation of Outcome	Six-Year Follow-Up Survey Question(s)
Exploratory Outcomes		
Number of school performance-related risks	Number of school-related risks perceived by parent for focal child (academic risk, attendance risk, and behavior risk), ranging from 0 to 3. Scale reflects the number of domains where risk is present (as evidenced by either of 2 statements being true): *Academic Risk* • Child has repeated any grades in school • Teacher has contacted an adult in the household this school year about any problems with schoolwork *Attendance Risk* • Child was absent from school for more than 2 days in the last month for any reason • Child was late for school on more than 2 days in the last month *Behavior Risk* • Teacher has contacted an adult in the household this school year about any behavior problems in school • Child has been suspended or expelled from school in the current school year	H16-H21

Outcome	Details on Derivation of Outcome	Six-Year Follow-Up Survey Question(s)
Parent believes focal child will graduate college	Parent believes child will finish college or earn an advanced degree after college.	H9
Parent is highly engaged	This scale was developed for the 3-year evaluations of PACE and HPOG 1.0. It was based on imputed average hours of time per day spent with the child in the typical week. The algorithm was different for preschoolers versus school-age children. Both thresholds were set at the 75th percentile for all children in the pooled evaluation samples for PACE and HPOG 1.0. For preschoolers, parents were credited with 1 hour for each shared breakfast in the typical week and 1 hour for each shared dinner. These hours were summed and then divided by 7. If the quotient was greater than the 75th percentile the parent was said to be highly engaged with the preschooler. For school-age children, parents were credited with 1 hour for each shared breakfast in the typical week, 1 hour for each shared dinner, 7 hours if they were usually present before the child leaves for school, 7 hours if they were usually present after the child comes home from school, 7 hours if they were usually present after dinner, 7 hours if they were present with the child during the weekend, and 7 hours if they talk to their child about homework during the week. These hours were summed and then divided by 7. If the quotient was greater than the 75th percentile, the parent was said to be highly engaged with the school-age child.	H4, H4a, H5, H5a, H6, H6a, H7, H7a, H8a, H8b, H11

B.2 Classification of Earned Credentials – Issuer and Required Length of Study

The six-year survey asked respondents about all credentials earned since the prior survey interview for people interviewed at three years, and all credentials earned since randomization for people not interviewed at three years. The questions closely paralleled those asked at three years. The instrument was based on several understandings about the nature and variety of credentials other than college degrees.

Consistent with the recommendation of a federal taskforce,[23] the instrument made a major distinction between "seat-time" credentials and "exam-based" credentials, where seat-time credentials are awarded by schools and other training providers to people who successfully complete required classes, and exam-based credentials are awarded by other authorities such as state and local agencies, unions, professional associations, and companies to people who successfully demonstrate proficiency at required skills, usually through passing an exam. In these reports, the first set of (seat-time) credentials are generally called *certificates, diplomas,* or *degrees.* The second set (exam-based) are generally called *certifications* and *licenses*. Note

[23] The task was formed in 2009 and formally known as the Interagency Working Group on Expanded Measures of Enrollment and Attainment (GEMEnA). Additional information on their work can be found at: [Adding-questions-on-certifications-and-licenses-to-the-current-population-survey.pdf (bls.gov)]

that this terminology implies an important distinction is made between *certificates* (which are awarded for seat-time) and *certifications* (which are awarded based on demonstration of skills).

The evaluation's analysis plan (Abt Associates 2015) requires two types of classification for each earned seat-time credential:

- *Is the issuer a college or some other training provider?*—This distinction is motivated by the conjecture that training at "colleges"[24] is intrinsically more valuable than training at other postsecondary institutions. This conjecture is based on the observation that college systems often offer to build and transfer stackable credentials—including potential for progressing from non-credit training to higher levels of credit-based instruction and credentials. Moreover, some of the PACE programs had the explicit goal of the earning of college-issued credentials.
- *How long are students typically required to study to earn the credential?*—The motivation for this is some econometric evidence that credentials that require longer training are likely to be more valuable on the job market. (See e.g., Dadgar and Trimble [2015]; Stevens, Kurlaender and Grosz [2019].)

Certain features of the survey instrument make it difficult to classify some credentials. Section B.2.1 documents procedures for classifying issuers. Section B.2.2 documents procedures for classifying the required length of study for earned seat-time credentials.

B.2.1 Classifying Issuers of Seat-Time Credentials

Analysis of the 18-month follow-up survey revealed that respondents found it quite difficult to report the nature of the institution at which they received training. To address this difficulty, the instrument used for both the three- and six-year surveys does not have a direct question on this. Instead, people are simply asked to report the name of the institution that issued their credential. As interviewers type the response into their laptop, a list of school names from the Department of Education's Integrated Postsecondary Education Data System (IPEDS) that match the first few typed characters appears. The interviewer can either select a name from the IPEDS list or type in the respondent-reported name verbatim.

The research team then used a combination of a Bayesian matching system and clerical review either to match the issuer name to a school name in IPEDS or to declare that the issuer name matches a non-IPEDS institution with a public web presence.[25] For issuers with names matched to an IPEDS entry, the study used the Carnegie classification of the school in IPEDS to classify the issuer as either a four-year college, a two-year college, or a non-degree-granting Title IV

24 The term "college" has no universal definition and there is no legal authority governing its usage. Our preferred definition is that embodied in IPEDS—namely degree-granting postsecondary institutions eligible to participate in federal Title IV financial aid programs.

25 One question about credentials asks the respondent whether they earned the credential by taking regular college courses. Though it would have been possible to assume that all issuers of these credentials were colleges, it was obvious that many of the issuers named by respondents were not, in fact, colleges. We did not make this assumption or otherwise use this information in classifying the provider as a college or some other type of training provider.

postsecondary school. For issuers declared not to be in IPEDS, the issuer is recoded as “other training provider.” This category includes non-Title IV schools (such as beauty and barber schools), employers, and social service agencies.

This section now provides more information on the Bayesian matching system and clerical review. Respondents and interviewers frequently misspelled the names of training providers. We used a combination of natural language processing (explained below) and clerical review to match “verbatim” names of providers with standardized spellings of the same. Classification was then automatic.

Natural Language Processing

Natural Language Processing (NLP) is a branch of machine learning dedicated to the analysis of natural languages.[26] At its core, NLP involves breaking text into distinct units, called *tokens*, and performing statistical analysis on the tokens. The process of breaking down the input text, called *tokenization*, as well as the types of statistical analysis employed vary with the NLP technique. For the purposes of credential and school name matching, we used two different tools: Naïve Bayes classification and fuzzy string matching.

Naïve Bayes Classification. Correction of misspellings can be viewed as a matching task, wherein verbatim names are matched to standardized names. It starts with an accepted set of matches (made in our case by clerical review of school/credential names across projects at Abt Associates), attempts to discover the subconscious rules that humans employ in manual matching, and then applying these “discovered” rules to match new verbatim responses.

Naïve Bayes classification uses a “bag of words” approach to classify strings of text.[27] Using bag of words, a string is broken down into words without regard for order or structure, where blank spaces and punctuation separate words. Consequently, acronyms are treated as individual words.[28] In this setting, manually coded data are represented as a count matrix, where each unique word that appears in the data corresponds to a row of the matrix, and each “document” (here the verbatim school and credential names) in the data corresponds to a column. Prior to assembling such a matrix, we first converted all the words in the data to lowercase to avoid case sensitivity.

At this point, the analysis has a count matrix where each word/document pair corresponds to the number of times the word appears in the document. The analysis then employs a “term frequency-inverse document frequency” transformation to account for how often given words appear in each document and in the “corpus” (the collection of all documents) as a whole. This

---

26 “Natural language” refers to a language that developed naturally (e.g., English), as opposed to an artificial language (e.g., Python™).

27 Mosteller and Wallace (1963) famously used a Bayesian framework for text classification to solve the puzzle of the “disputed” Federalist Papers, 12 papers for which historians did not agree upon the true author. They showed convincing evidence that it was James Madison.

28 We expect many acronyms in the school name matching. Though treating acronyms as individual words may not appear ideal, if an acronym is matched with a school in the training set, it will help guide the naïve Bayes algorithm to match future instances of that acronym with the school.

transformation, defined by Equations (B.2.1a) and (B.2.1b) below, helps control for the size of the documents and the relative frequency of words. For example, without this transformation, larger documents with more words might overpower smaller documents in the analysis. Additionally, controlling for the relative frequency of words in the corpus limits the influence of frequently occurring words.[29]

$$TF_{t,d} = \frac{F_{t,d}}{NT_d} \qquad \text{(Equation B.2.1a)}$$

$$IDF_t = \frac{M}{df_t} \qquad \text{(Equation B.2.1b)}$$

where t, d index the word in question and the document in question; $F_{t,d}$ is the frequency of word t in document d; NT_d is the number of words in document d; M is the number of documents in the corpus; and df_t is the number of documents in the corpus that contain word t.

In the transformed matrix, called the *term frequency-inverse document frequency matrix*, the entry for row t, column d is the product $TF_{t,d} \times IDF_t$.

Once we have transformed the data as described above, we can implement the naïve Bayes algorithm to help match survey responses to standardized names. Suppose we have a set of m standardized names and a verbatim survey response comprising words $x_1, \ldots, x_n$. Considering binary indicator variables $y_1, \ldots, y_m$, where $y_j = 1$ if the document belongs to class j (i.e., the verbatim string is an alternate spelling of the indicated standardized name), Bayes' Theorem gives us the following likelihood that the verbatim response truly corresponds to standardized name j:[30]

$$P(y_j | x_1, \ldots, x_n) = \frac{P(y_j) P(x_1, \ldots, x_n | y_j)}{P(x_1, \ldots, x_n)} \qquad \text{(Equation B.2.1c)}$$

Additionally, we make the naïve conditional independence assumption that the probabilities $P(x_i | y_j)$ are independent given class j; that is, for all $i = 1, \ldots, n$ we have:

$$P(x_i | y_j, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = P(x_i | y_j) \qquad \text{(Equation B.2.1d)}$$

In other words, given that a string belongs to class j, we assume the likelihood that we see a given word is independent of the other words in the string. This assumption may not always hold up in the real world. For example, suppose that a given survey response for a school name truly corresponds with "University of Georgia." If the verbatim survey response contains the word "Univ," the respondent/survey administrator is likely to have abbreviated other words as well, so it may be more likely that "GA" will also be in the response than in other circumstances. Though

29 For example, in the school matching, we would expect the word *college* to appear in many of the reported names as well as the standard names. In this case, we would not want the presence of the word *college* to be highly significant in an individual matching.

30 As is common convention, we use $P(y_j)$ as a shorthand for $P(y_j = 1)$.

this assumption may seem overly simplistic, naïve Bayes classifiers perform quite well in real-world situations.[31]

The naïve conditional independence assumption simplifies Equation (C.2.1c) to:

$$P(y_j|x_1,\dots,x_n) = \frac{P(y_j)\prod_{i=1}^{n} P(x_i|y_j)}{P(x_1,\dots,x_n)} \qquad \text{(Equation B.2.1e)}$$

$P(x_1,\dots,x_n)$ is constant given the input; therefore, we match the verbatim survey response with standardized name $\hat{j}$, where:[32]

$$\hat{j} = \underset{j=1,\dots,m}{\arg\max}\, P(y_j)\prod_{i=1}^{n} P(x_i|y_j) \qquad \text{(Equation B.2.1f)}$$

We can then use maximum a posteriori estimation to estimate $P(y_j)$ and $P(x_i|y_j)$.[33] The estimate for $P(y_j)$ then becomes the relative frequency of the class j in the training set.[34] The assumed distribution of $P(x_i|y_j)$—that is, the probability a document in class j contains word x_i—depends on the exact naïve Bayes classifier used. For this evaluation, we fit a multinomial naïve Bayes classifier using the Scikit-Learn Python package.[35]

Here we use the following estimator, $\hat{\theta}_{ji}$, for $P(x_i|y_j)$:[36]

$$\hat{\theta}_{ji} = \frac{N_{ji}+1}{N_j+n} \qquad \text{(Equation B.2.1g)}$$

where N_{ji} is the number of times x_i has appeared in a sample of class j in the training set, weighted by term frequency-inverse document frequency; that is, $N_{ji} = \sum_{d\in y_j} TF_{x_i,d}\ x\ IDF_{x_i}$.[37] Lastly, $N_j = \sum_{i=1}^{n} N_{ji}$.

Fuzzy String Matching. In this report, we fit a fuzzy string matching model using a vectorial decomposition approach through the Stata *matchit* command. Fuzzy matching allows strings from one set to be matched to strings from a second set, where similarity scores define string-wise matchings. In such a model, text strings are broken down into elements of two characters (called *bigrams*), and the similarity of two strings depends on the number of common bigrams

---

[31] For discussions on real-world applications of naïve Bayesian techniques, see Zhang (2004) and Chen et al. (2016).

[32] Recall that n is the number of words in the verbatim survey response and m is the number of standard classifications.

[33] *Maximum a posteriori estimation* refers to the Bayesian technique of using the mode of the posterior distribution of a particular parameter as an estimate for that parameter.

[34] For example, the proportion of students in the training set who went to school j.

[35] Pedregosa et al. *(*2011).

[36] Rennie et al. (2003).

[37] In other words, N_{ji} is the sum of the row corresponding to term x_i in the term frequency-inverse document frequency matrix restricted to columns corresponding to documents in the class y_j.

they share.[38] For example, if we wished to compare "Smith, John" with "Smit, John," we would proceed as shown in Exhibit B-7.

Exhibit B-7: Example of Fuzzy String Matching

String	Element Number									
	1	2	3	4	5	6	7	8	9	10
smith, john	**sm**	**mi**	**it**	th	h,	**,**	**j**	**jo**	**oh**	**hn**
smit, john	**sm**	**mi**	**it**	t,	**,**	**j**	**jo**	**oh**	**hn**	

Prior to decomposition, we converted strings to lowercase to avoid case sensitivity. The bolded bigrams above are shared between the two strings. The strings have 10 and 9 bigrams, respectively, and share 8. Using this data, we can then calculate a similarity score between strings i (Str_i) and j (Str_j) using the formula described in Equation B.2.1h:[39]

$$Similarity(Str_i, Str_j) = \frac{m}{\sqrt{s_i s_j}} \qquad \text{(Equation B.2.1h)}$$

where s_i and s_j are the number of bigrams in strings i and j (respectively) and m is the number of matching bigrams in the two strings. For example, the similarity for the two scores above would be $\frac{8}{\sqrt{90}} \sim 0.84$. A similarity score of 1 corresponds to a perfect match, whereas a score of 0 corresponds to no common elements. When matching strings from Set A with strings from Set B, a given string from Set A is matched with whichever string from Set B gives the highest similarity score.[40] In other words, given Str_a in Set A, we use the following formula to determine $match_{Str_a}$, its counterpart in Set B:

$$match_{Str_a} = \underset{Str_b \in Set\ B}{\operatorname{argmax}}\ Similarity\ (Str_a, Str_b) \qquad \text{(Equation B.2.1i)}$$

School Matching

We implemented a dual naïve Bayes/fuzzy matching approach to match survey-reported school names to standardized names.

This approach was a seven-step process in which we[41]

(1) created an initial list of standardized issuer names by compiling a list of schools reported by participants in previous Abt analyses;

(2) matched the survey-reported names to schools in this initial standardized list using the fuzzy matching algorithm;

---

38 This technique can be applied to elements of any number of characters.

39 There are multiple approaches to calculating similarity scores. This is known as a Jaccard similarity score.

40 Raffo (2020).

41 Survey administrators had the ability to select IPEDS schools from a drop-down menu while conducting the survey. Of all reported credential issuers, 50 percent were selected in that manner and were not subject to this matching algorithm.

(3) manually reviewed potentially problematic matches from the list of proposed matches generated above and rejected proposed matches that did not appear valid;

(4) joined the list of all IPEDS schools with the standardized list described above to create a new list of standardized issuer names and a training set for the naïve Bayes algorithm;[42]

(5) ran the remaining unmatched survey responses through the fuzzy matching algorithm (against the new list of standardized issuer names described in Step 5), as well as through a naïve Bayes model trained on the aforementioned training set to produce two sets of proposed matches;

(6) reviewed both sets of proposed matches; in the cases where the two algorithms differed with respect to their proposed name, at most one name could be accepted;[43] and

(7) manually searched for schools in the remaining list of rejected matches (475 survey responses fell into this category).

B.2.2 Classifying Required Length of Study for Seat-Time Credentials

The instruments for the three- and six-year surveys asked about the typical required length of study for some but not all certificates and diplomas. As noted in the first paragraph of Section B.2, the instrument made a major distinction between certificates/diplomas that involved taking "regular college classes" and those that did not. For those that did involve taking regular college classes, the instrument asked about the required length of study; for other certificates, the plan was to code the required length of study in post-processing, based on the respondent-reported name of the certificate. This plan was followed in the three-year reports for all PACE sites. However, further exploration suggested that while the imputation procedure worked well for healthcare certificates, it did not work well for certificates in the field of information technology. The names reported by respondents were simply too ambiguous. Moreover, responding to interviewer reports about respondent confusion about how to report credentials according to this schema in the three-year survey, we made some wording changes to the instrument for the six-

[42] We created the training set by creating a list of synthetic matches (i.e., matching standardized names to themselves) using this updated list of standardized names. We also included matches from the earlier fuzzy matching phase and previous Abt analyses. Including the synthetic matches ensured that each standardized name is in the training set (and is thus an eligible category for matching) and helps guide the naïve Bayes matching algorithm by promoting matches between survey responses and standardized names with similar words. Returning to the previous example of the University of Georgia: If the only previous survey responses referring to this school had referred to it by its common abbreviation, "UGA," then a naïve Bayes algorithm trained on those examples would not be equipped to match a survey response of "University of Georgia" to the standardized school of the same name. Including the synthetic matches helps avoid this issue.

[43] We attempted to match 689 survey responses at this step. The naïve Bayes and fuzzy matching algorithms agreed in 144 cases (21 percent). We accepted 126 of these 144 proposed matches, a substantially higher rate than when the two algorithms disagreed. Among the 545 cases where the two algorithms disagreed, we accepted the naïve Bayes match 13 percent of the time and the fuzzy match 3 percent of the time.

year survey. These changes seemed to lead respondents to classify more credentials as certificates not involving the taking of regular college classes. As a result, the number of certificates with an unknown required length of study increased strongly from the three-year survey to the six-year survey.

To deal with these issues, the team developed a new classification of seat-time credentials different than envisioned in the analysis plan. The new hierarchy is as follows:

1. A college-issued degree (according to either the survey or the NSC).
2. Some other college-issued certificate or diploma.
3. Any certificate or diploma issued by some training provider other than a college.

We applied this hierarchy retroactively to credentials reported both in the three-year survey and in the six-year survey, dropping the earlier classification system for the earlier reported credentials. Not being able to classify the required length of study for certificates awarded by training providers other than colleges probably is not a serious problem because trainings that require at least a year of study are more likely to be delivered at colleges than at other training providers.

B.3 Imputation in the Six-Year Survey

As in any survey, some respondents do not answer every question. Rather than dropping respondents with missing survey items, we used a variety of approaches to make use of the partial responses. Our decision to include or drop such cases varied, depending on the frequency of nonresponse to the question across respondents.

The default rule was to drop persons from any analysis involving unanswered questions but to include them for all other analyses. Where this rule would result in a sharp drop in sample size, we instead imputed responses for those people for those questions, rather than dropping them.

The goals of imputation were variance and bias reduction.[44] Both goals are achievable with the rich set of parallel outcomes measured in the six-year survey. For example, indications of problems paying bills is valuable information for imputing missing income. Specifically, we imputed for seven types of missing data:

(1) Credential award date.

(2) Level of school for credential issuers (four-year college, at least two- but less than four-year college, less than two-year, not Title IV postsecondary institution).

(3) Early certificates and licenses (first 18 months after randomization).

---

[44] Systematic nonresponse (e.g., those without college credentials are less likely to answer questions about credential attainment) can cause biased estimates. Effective imputation can reduce this bias. Making use of more data also increases sample size, thereby reducing the variance of impact estimates.

(4) Income (personal and household).

(5) Student debt (in own name and parent's name) for Year Up only.

(6) Enrolled in education or training in the last three years.

(7) Exact amount of "other debt" given respondent-provided information on bracketed amounts of such debt.

Section B.3.1 briefly describes the prevalence of missing data along with each type of imputation. With the exception of level of school (2) and the exact amount of other debt (7), we used a common methodology for all types of missing data. Section B.3.2 provides the detail on these imputations. For the most part, imputation only utilized information from within the same program as the one with missing data. However, there are steps in the imputation process where the imputation did use cross-program information, including information from the parallel evaluation of HPOG 1.0 since the evaluations used nearly identical instruments and were fielded about the same time. These instances are clearly noted below. Appendix E looks at the sensitivity of impacts to imputation.

B.3.1 Missing Item Rates

Exhibit B-8 below lists the first six types of imputation and shows the imputation rates for the survey respondents in the four PACE sites with a six-year follow-up survey. (The imputation of exact other debt given bracketed debt is not shown as it was universal.) Respondents had trouble recalling the dates on which they received credentials. Income was also frequently missing, especially household income. The instrument prompted respondents to give a categorical answer ("bracketing") if they could not give an exact figure.

B.3.2 Imputation Procedure

As mentioned above, five of the seven types of imputation used a common imputation procedure. This section discusses the procedure used and provides additional details for each of the seven types of missing data.

Core Imputation Procedure. The core imputation methodology involved four steps. The first step entailed assembling a list of potential predictors and imputing any missing data in them.[45] The list of potential predictors included treatment status, baseline variables, parallel outcomes, and two-way and three-way interactions of both baseline variables and parallel outcomes with treatment status.

---

[45] The only purpose of the imputation of potential predictors was to facilitate automated variable selection in the next step. After we used these imputed values of the predictors to predict new exam-based certifications and licenses as of the time of the short-term survey, we discarded them. We carried out this imputation with SAS/MI/FCS.

Exhibit B-8: Imputation Rates among Survey Respondents in Four PACE Sites, Six-Year Survey

Type of Imputation	CES Credentials (%)	CES People (%)	I-BEST Credentials (%)	I-BEST People (%)	VIDA Credentials (%)	VIDA People (%)	Year Up Credentials (%)	Year Up People (%)
1. Credential award date	9.2	n/a	2.6	n/a	5.0	n/a	4.6	n/a
2. Level of school (credential issuer)	6.6	n/a	10.0	n/a	2.3	n/a	9.2	n/a
3. Early certifications and licenses (new imputations at six years)	n/a	2.1	n/a	7.0	n/a	4.6	n/a	5.8
3. Early certifications and licenses (imputed at three years)	n/a	6.0	n/a	8.4	n/a	5.6	n/a	8.5
4. Income								
Personal (categorical)	n/a	1.5	n/a	1.1	n/a	3.4	n/a	2.2
Personal (exact)	n/a	5.4	n/a	2.2	n/a	5.1	n/a	6.5
Household (categorical)	n/a	4.5	n/a	6.1	n/a	6.7	n/a	5.8
Household (exact)	n/a	18.7	n/a	9.8	n/a	13.0	n/a	24.5
5. Student debt								
Own (categorical)	n/a	0.2	n/a	0.8	n/a	0.4	n/a	1.4
Own (exact)	n/a	0.7	n/a	0.8	n/a	1.2	n/a	3.2
Parents (categorical)	n/a	0.4	n/a	0.0	n/a	0.3	n/a	1.0
Parents (exact)	n/a	0.4	n/a	0.0	n/a	0.4	n/a	1.5
6. Enrolled in education/training in last three years	n/a	3.9	n/a	7.5	n/a	7.4	n/a	5.8

Key: n/a = not applicable.

Source: PACE six-year follow-up survey.

Note: Exact income was missing more often than categorical income because respondents unable or unwilling to provide an exact amount were encouraged to report a bracketed amount. The imputation of exact other debt given bracketed debt (7) is not shown as it was universal.

The second step entailed the use of the least absolute shrinkage and selection operator (LASSO) procedure to fit a linear model for the target variable in terms of the assembled predictor list.[46] We ran the LASSO separately by PACE site unless otherwise noted.

The third step used predicted values from the final linear model to create a nested set of three partitions for each treatment status and PACE site.[47] The finest partition involved splitting the sample into 20 equal-sized groups based on the predicted probability of having reported the outcome. The middle partition corresponded to deciles of this same probability, and the coarsest partition corresponded to quintiles of this same probability.

The fourth and final step used the hotdeck imputation procedure in SUDAAN to randomly match each nonrespondent with a respondent within cells defined by treatment status and PACE site and the nested partitions. Most cases were matched within cells defined by the 20-level partition. If there were no matches within those cells, then the procedure sought matches within the coarser partitions.

We ran the final hotdeck procedure five times with different random seeds to produce multiple imputations. We used these multiple imputations in the formal analysis runs to add between-imputation variance onto the naïve variance estimates on the full sample, using Rubin's classic formula.[48]

Summarized Well-Being. The survey contained multiple measures of financial and social-emotional well-being. We theorized that these variables would be useful predictors of several types of missing data. However, interpretation of high-dimensional models is difficult. As a way of incorporating these rich data on well-being into imputation models while keeping the models fairly easy to interpret, we condensed all these measures into a partition of the sample using cluster analysis. We were able to identify five clusters of respondents in the pooled PACE and HPOG 1.0 sample that vary clearly in terms of quality of life, career progress, and family dependence. For shorthand, we refer to them as "life trajectory" clusters because one of the variables that they vary on most clearly is a sense of career progress:

- "Overextended"—above-average income but also above-average financial stress and low scores on psycho-social skills.
- "Family supported"—below-average income but strong family supports that protect them from financial stress.
- "Strivers"—strong psycho-social skills and sense of career progress but low income (personal and household) and dependent on public support.
- "Down and out"—very low psycho-social skills, low sense of career progress, severe life challenges, low income (personal and household), and strong reliance on public support.

46 See Appendix Section A.2 for details on the cross-validated LASSO.

47 A "partition" of a sample is an exhaustive and mutually exclusive collection of subsets of the sample.

48 See for example, Rubin (1987).

- "Winners"—strong psycho-social skills and sense of career progress, high income (personal and household), few bill problems, and little dependence on either family or public support.

(1) Missing Credential Award Dates

We imputed missing credential award dates by imputing the lag between randomization date and credential award date and then adding that imputed lag onto the actual randomization date for the person. This imputation of lags followed the core procedure described above. Specifically, we modeled the lag between randomization and credential award date for those respondents with reported award dates (*n*=2,475, with 2,342 responses). We pooled credentials across all PACE sites and HPOG 1.0. The potential predictor list included site, treatment, the interaction of site and treatment, an indicator of whether the interview took place after March 1, 2020,[49] type of credential (three categories), type of degree, life trajectory cluster, 17 parallel outcomes at the person-level (e.g., perceived progress toward career goals), the lag between randomization and interview, and 16 baseline variables. After creating dummy variables for categorical variables, the total number of potential predictors was 67.

The LASSO procedure working on this predictor set selected six variables:

- Two dummy variables for sites Carreras en Salud and Year Up.
- A dummy variable for credential type (government/industry certification/license).
- A dummy variable for obtaining a bachelor's degree or higher.
- A dummy variable for the overextended life trajectory cluster.
- A dummy variable indicating whether the participant's age was between 25 and 34 at baseline.

The R-squared was 2.2 percent.[50]

After matching nonrespondents with respondents, we adjusted for the difference in randomization dates between the two people, by adding the lag from the respondent to the randomization date for the nonrespondent. If this was past the interview date for the nonrespondent, we truncated the award date to equal the interview date.

(2) Missing School Level for Credential Issuers

On the PACE credential sample, after we classified issuers of credentials (see Section B.2.1), we imputed missing school level for credential issuers (whether the issuer was a four-year college, a college offering at least two-year degrees but less than four-year degrees, a school

49 We included this indicator in all of the imputations to account for differences that may have occurred after COVID-related shutdowns.

50 This low value indicates that the imputation is unlikely to reduce the true variance on the estimated impacts of PACE programs on the lag between randomization and award date, but the imputation is nonetheless important because we need a date of some sort to deduplicate credentials reported in the three-year and six-year follow-up surveys.

offering credentials of less than two years, or not a Title IV postsecondary institution). About one third of the credentials missing school level were due to respondents not remembering the name of the credential issuer, and two thirds were due to issuers that were difficult to classify. We imputed issuer level using treatment, site, type of credential (academic, vocational, or certification/license) and type of degree (diploma requiring less than a full year's worth of credit, diploma requiring a full year or more but less than an associate degree, an associate degree, or a bachelor's degree or higher). We pooled PACE and HPOG 1.0 credentials and used single imputation using the SUDAAN imputation procedure instead of the core imputation procedure described above. We used treatment status, site, type of credential (academic, vocational, certification/license) and degree type to determine the hard cells.

(3) Certifications and Licenses in the First 18 Months

We imputed early certifications and licenses for study participants who were not interviewed at 18 months after randomization or at three years but who were interviewed at six years. Earlier analyses identified an issue with the quality of reports of receipt of exam-based credentials in the three-year follow-up survey, which was corrected by using data on these credentials from the 18-month survey, imputing the missing credentials first for people who skipped the 18-month interview.[51] As part of the work on the three-year reports for PACE programs, we had already imputed the missing 18-month data for people who skipped the 18-month survey interview but completed the three-year survey. For this round of reports, we had to also impute the missing 18-month data for people who skipped both the 18-month and the three-year surveys.[52] We used the core imputation described above for this imputation.

We modeled the receipt of these credentials among those who responded to the 18-month survey by PACE site. The potential predictor list included Year Up program office (for the Year Up model only) and I-BEST campus (for the I-BEST model only), treatment status, and about 30 baseline and six-year follow-up variables. After creating dummy variables for levels of categorical variables, this led to 37 potential predictors in total.

The **Carreras en Salud** six-year survey respondent sample included 536 people, of whom 493 responded to both the 18-month survey and the six-year survey. The LASSO selected the following variables as predictors:

- Treatment status.
- Receipt of a college certificate requiring less than one year of study.
- Working 35 or more hours per week at baseline.

---

51 See Section B.4 of Judkins et al. (2020).

52 Nonrespondents here were people who could not be located, refused to be interviewed, or were otherwise unavailable for a survey interview. The concept does not include people who skipped questions about credentials when interviewed at 18 months. We assumed that these respondents did not earn any credentials by the time of the 18-month interview.

The *R*-squared was 7.1 percent.[53]

The **I-BEST** six-year survey respondent sample included 358 people, of whom 295 responded to both the 18-month survey and the six-year survey. The LASSO selected the following variables as predictors:

- Treatment status.
- The Whatcom campus.
- The interaction of treatment status with the Everett campus.
- Gender.
- Receipt of a college certificate requiring less than one year of study.

The *R*-squared was 15.1 percent.

The **VIDA** six-year survey respondent sample included 732 people, of whom 657 responded to both the 18-month survey and the six-year survey. The LASSO selected the following variables as predictors:

- A dummy variable indicating reliance on public support.
- A dummy variable indicating whether the participant's age was between 21 and 24 at baseline.
- Number of certifications/licenses received.
- Receipt of a college certificate requiring less than one year of study.
- Receipt of a college certificate requiring more than one year of study but less than an associate degree.
- Receipt of an associate degree.

The *R*-squared was 8.2 percent.

The **Year Up** six-year survey respondent sample included 1,653 people, of whom 1,410 responded to both the 18-month survey and the six-year survey. The LASSO selected the following variables as predictors:

- Treatment status.
- Interaction of treatment status with one of the Year Up offices.
- Working 35 or more hours per week at baseline.
- Number of academic credentials received.
- Number of certifications/licenses received.
- Receipt of a college certificate requiring less than one year of study.

---

53 We fit linear models for the prediction phase, even for binary outcomes. We did this because the linear LASSO is much faster than the logistic LASSO.

- Receipt of a college certificate requiring more than one year of study but less than an associate degree.

The *R*-squared was 6.2 percent.

For the hotdeck step, 500 records were imputed at the 20-level partition and 1 record was imputed at the 10-level partition. No records were imputed at the 5-level partition.

After imputing exam-based certifications and licenses for short-term survey nonrespondents, we separated exam-based certifications and licenses reported in the six-year survey using the donor's interview date into two categories—early (would have been reported by the nonrespondent in 18-month survey if the interview had taken place) versus late (would have been earned after the survey if the interview had taken place). We then created a blended flag for having earned an exam-based certification or license as of the six-year survey as follows:

- For six-year respondents who responded at 18 months and three years, we kept the composite measure created at three years and added receipt of exam-based credentials between three years and the six-year survey.
- For six-year respondents who did not respond at 18 months but responded at three years, we kept the imputation done at three years and added receipt of exam-based credentials between three years and the six-year survey.
- For six-year respondents who responded at 18 months but did not respond at three years, we created a composite measure of receipt of any exam-based credential since randomization, which was set to yes if the respondent reported either an exam-based credential in the 18-month survey or in the six-year survey at a point in time after the 18-month survey interview date.
- For six-year respondents who did not respond at 18 months or three years, we used the imputed response for early certifications and licenses using the procedure described above and combined it with exam-based credentials earned between 18 months and six years.

(4) Missing Income

The instrument yielded four related measures of income in the past month: (1) exact personal income (as best recalled); (2) categorical personal income (from a bracketed response for respondents who were unable or unwilling to provide an exact amount); (3) exact household income (as best recalled); and (4) categorical household income (again, from a bracketed response for respondents who were unable or unwilling to provide an exact amount). As seen in Exhibit B-8, missing data rates were higher for the continuous variables than for the categorical variables. This is true by construction. Missing data rates are coded such that categorical income is only missing if both exact income (which can be put in the appropriate income category) and categorical income are missing.

For prediction purposes, we assembled a person-level file with treatment status, Year Up office (Year Up model only), I-BEST campus (I-BEST model only), seven variables about economic well-being, one variable about goal progress, seven measures of educational progress, 12 baseline characteristics, one variable about family structure, the life trajectory cluster described earlier, an indicator of whether the interview took place after the start of the COVID-19 pandemic (operationalized as March 1, 2020), and personal and household income from the three-year survey. We used this list to model both personal and household income. We ran the LASSO on the six-year survey data by PACE site. After creating dummy variables for categorical variables, the total number of potential predictors was 40.

The **Carreras en Salud** six-year survey respondent sample included 536 people, with 507 exact personal income reports and 436 exact household income reports. The LASSO selected the following variables as predictors for personal income:

- Dummy variables for the family supported, striver, and winner life trajectory clusters.
- A dummy variable indicating reliance on public support.
- Receipt of a college certificate requiring more than one year of study but less than an associate degree.
- Receipt of an associate degree.
- Personal income at the three-year survey.

The LASSO selected the following for household income:

- Dummy variables for the down and out, striver, and winner life trajectory clusters.
- A measure for the amount of money left over at the end of the month.
- A dummy variable indicating reliance on public support.
- A dummy variable indicating whether the respondent claimed the Earned Income Tax Credit (EITC) last year.
- Number of academic credentials received.
- Household income at the three-year survey.

The *R*-squared was 5.5 percent for personal income and 3.6 percent for household income.

The **I-BEST** six-year survey respondent sample included 358 people, with 350 exact personal income reports and 323 exact household income reports. The LASSO selected the following variables as predictors for personal income:

- Dummy variables for the family supported, overextended, striver, and winner life trajectory clusters.
- Gender.
- Personal income at the three-year survey.

The LASSO selected the following for household income:

- Dummy variables for the down and out, striver, and winner life trajectory clusters.
- A measure for the amount of money left over at the end of the month.
- A dummy variable indicating reliance on public support.

- Household income at the three-year survey.

The *R*-squared was 5.0 percent for personal income and 4.5 percent for household income.

The **VIDA** six-year survey respondent sample included 732 people, with 695 exact personal income reports and 637 exact household income reports. The LASSO selected the following variables as predictors for personal income:

- Dummy variables for the down and out, family supported, striver, and winner life trajectory clusters.
- A measure for the amount of money left over at the end of the month.
- A dummy variable indicating reliance on public support.
- A dummy variable indicating whether the respondent claimed the EITC last year.
- Number of certifications/licenses received.
- Receipt of an associate degree.
- Personal income at the three-year survey.

The LASSO selected the following for household income:

- Dummy variables for the striver and winner life trajectory clusters.
- A measure of food insecurity.
- A measure for the amount of money left over at the end of the month.
- A dummy variable indicating reliance on public support.
- A dummy variable indicating whether the respondent claimed the EITC last year.
- Number of certifications/licenses received.
- Household income at the three-year survey.

The *R*-squared was 5.4 percent for personal income and 5.0 percent for household income.

The **Year Up** six-year survey respondent sample included 1,653 people, with 1,546 exact personal income reports and 1,248 exact household income reports. The LASSO selected the following variables as predictors for personal income:

- Treatment status.
- The Atlanta, Boston, and NCR Year Up offices.
- The interaction of treatment status with the Boston and NCR Year Up offices.
- A dummy variable indicating that the interview took place after March 1, 2020.
- Dummy variables for the down and out, family supported, striver, and winner life trajectory clusters.
- Scale for progress toward goals.
- A measure for the amount of money left over at the end of the month.

- A dummy variable indicating whether the respondent was dependent on family.
- A dummy variable indicating reliance on public support.
- A dummy variable indicating whether the respondent claimed the EITC last year.
- A dummy variable indicating whether the participant's age was between 21 and 24 at baseline.
- Personal income at the three-year survey.

The LASSO selected the following for household income:

- The NCR, New York City, and Puget Sound Year Up offices.
- The interaction of treatment status with the Atlanta Year Up office.
- Dummy variables for the striver and winner life trajectory clusters.
- A measure for the amount of money left over at the end of the month.
- A dummy variable indicating reliance on public support.
- A dummy variable indicating whether the respondent claimed the EITC last year.
- Personal income at the three-year survey.
- Household income at the three-year survey.

The R-squared was 5.2 percent for personal income and 4.4 percent for household income.

For the hotdeck steps for categorical household and personal income, all records were imputed at the 20-level partition. For the hotdeck step for exact household income, 880 records were imputed at the 20-level partition and 1 record was imputed at the 10-level partition. No records were imputed at the 5-level partition. For the hotdeck step for exact personal income, all records were imputed at the 20-level partition.

(5) Missing Student Debt

The instrument yielded four related measures of student debt: (1) exact personal student debt (as best recalled); (2) categorical personal student debt (if the respondent was unable or unwilling to provide an exact amount); (3) exact parental student debt (as best recalled); and (4) categorical parental student debt (if the respondent was unable or unwilling to provide an exact amount). As seen in Exhibit B-8, missing data rates were higher for the continuous variables than the categorical variables. This is true by construction given that categorical student debt is only missing if both exact student debt (which can be put in the appropriate debt category) and categorical student debt are missing. Missing rates for student debt were very low for all sites except for Year Up, so we used imputed values only for Year Up.

For prediction purposes, we assembled a person-level file with Year Up office, treatment status, seven variables about economic well-being, one variable about goal progress, seven measures of educational progress, highest level of educational attainment at the six-year survey, an indicator of whether the respondent's level of education was higher at six years than at baseline, 15 baseline characteristics, one variable about family structure, the life trajectory cluster described earlier, an indicator of whether the interview took place after March 1, 2020, personal and household income from the three-year survey, and four NSC variables on enrollment and credential attainment. We used this list for modeling both personal and parental student debt. We ran the LASSO on the Year Up six-year dataset (*n*=1,653, with 1,600 exact personal student debt reports and 1,628 exact parental student debt reports). After creating dummy variables for categorical variables, the total number of potential predictors was 63.

The LASSO selected the following variables as predictors for personal student debt:

- Treatment status.
- The Atlanta, Bay Area, and Puget Sound Year Up offices.
- The interaction of treatment status with the Atlanta and NCR Year Up offices.
- A dummy variable for the overextended life trajectory cluster.
- Scale for progress toward goals.
- A dummy variable indicating race is Black non-Hispanic.
- Receipt of a bachelor's degree.
- Highest level of educational attainment at the six-year survey.
- Full-time-equivalent (FTE) months of enrollment through Q24 from the NSC.

The LASSO did not select any variables as predictors for parental student debt. However, the hotdeck imputation step described earlier to randomly match each nonrespondent with a Year Up respondent within cells defined by treatment status and the nested partitions still worked.

The *R*-squared was 2.4 percent for personal student debt and 0 percent for parental student debt.

For the hotdeck steps for categorical and exact parental student debt and personal student debt, all records were imputed at the 20-level partition.

(6) Missing Enrollment in Education/Training in the Last Three Years

For three-year nonrespondents, we do not know whether they were enrolled in training in the last three years. Instead, we only know if they were enrolled since random assignment and if they are currently enrolled at the time of the six-year survey. We ran the LASSO on the six-year survey data by PACE site.

For prediction purposes, we assembled a person-level file with Year Up office (Year Up model only), I-BEST campus (I-BEST model only), treatment status, seven variables about economic well-being, one variable about goal progress, seven measures of educational progress, highest level of educational attainment at the six-year survey, an indicator of whether the respondent's level of education was higher at six years than at baseline, 15 baseline characteristics, one variable about family structure, an indicator for current wages at or above $15 per hour, the life trajectory cluster described earlier, an indicator of whether the interview took place after March 1, 2020, and four NSC variables on enrollment and credential attainment. After creating dummy variables for categorical variables, the total number of potential predictors was 52.

The **Carreras en Salud** six-year survey respondent sample included 536 people, with 506 reports of enrollment in education/training in the last three years. The LASSO selected the following variables as predictors:

- Scale for progress toward goals.
- A measure for the amount of money left over at the end of the month.
- Gender.
- Working 35 or more hours per week at baseline.
- Receipt of a college certificate requiring less than one year of study.
- Receipt of a college certificate requiring more than one year of study but less than an associate degree.
- Receipt of an associate degree.
- Three dummy variables for highest educational attainment at the six-year survey.
- A dummy variable indicating any enrollment in the NSC by Q24.
- Cumulative months of enrollment in the NSC by Q24.
- A dummy variable indicating that educational attainment was higher at the six-year survey than at baseline.

The *R*-squared was 4.1 percent.

The **I-BEST** six-year survey respondent sample included 358 people, with 328 reports of enrollment in education/training in the last three years. The LASSO selected the following variables as predictors:

- The interaction of treatment status with the Bellingham campus.
- Scale for progress toward goals.
- A measure for the amount of money left over at the end of the month.
- A dummy variable indicating whether the respondent was dependent on family.
- A dummy variable for Hispanic ethnicity.
- A dummy variable indicating race is Black non-Hispanic.
- A dummy variable indicating wages of $15 per hour or higher at the six-year survey.
- Two dummy variables for highest educational attainment at the six-year survey.
- Cumulative months of enrollment in the NSC by Q24.

The *R*-squared was 21.2 percent.

The **VIDA** six-year survey respondent sample included 732 people, with 667 reports of enrollment in education/training in the last three years. The LASSO selected the following variables as predictors:

- A dummy variable for the winner life trajectory cluster.

- Scale for progress toward goals.
- A measure for the amount of money left over at the end of the month.
- A dummy variable indicating reliance on public support.
- Number of vocational credentials received.
- A dummy variable indicating receipt of any credential in the NSC by Q24.
- Cumulative months of enrollment in the NSC by Q24.

The *R*-squared was 30.5 percent.

The **Year Up** six-year survey respondent sample included 1,653 people, with 1,512 reports of enrollment in education/training in the last three years. The LASSO selected the following variables as predictors:

- Scale for progress toward goals.
- Number of vocational credentials received.
- Number of certifications/licenses received.
- Cumulative months of enrollment in the NSC by Q24.

The *R*-squared was 23.5 percent.

For the hotdeck step, all records were imputed at the 20-level partition.

Exact Amount of Other Debt

To be able to calculate total unsecured debt (student debt plus other unsecured debt), it was necessary to impute the exact amount of unsecured other debt. We did this using the bracketed amounts reported by respondents as well as other information from the survey and an assumption about the shape of the true distribution of other debt.

Respondents were asked to report their other unsecured debt by brackets. These brackets were narrow for low levels of debt and increasingly broad for higher levels of debt. The broadest category was at the top ($50,000 or more). Except for this category, we might have merely imputed the midpoints of each bracket as the “exact” amount.[54] However, we believed that with about 2 percent of people reporting debt greater than $50,000, that it was worth doing something better.[55]

First, we imputed missing data for this categorical question using SAS PROC MI by site, based on study group (treatment versus control), any signs of financial distress, and whether the first scheduled survey interview attempt came after the outbreak of COVID-19. We then imputed

---

54 This is, in fact, what we did at three years, with a “middle” point for the top bracket of $75,000, following a common rule used at the Census Bureau to set the middle of unbounded income and asset brackets at 150 percent of the lower edge of the top bracket.

55 By site, the percentages reporting more than $50,000 in other unsecured debt were 1 percent for Carreras, 3 percent for I-BEST, 2 percent for VIDA, and 2 percent for Year Up.

exact amounts from the categorical responses. We did this by running SAS PROC LIFEREG separately by PACE site and treatment status.

The LIFEREG procedure assumes that among people with any unsecured debt, after controlling on known factors, the exact amount follows a log-normal distribution. As known factors we included personal income, categorical personal student debt, a flag for bill troubles, a scale of life challenges, whether they owned their own home, whether they lived with their parents, and whether they lived with a spouse and their own children. We chose these variables based on their predictive power on a pooled dataset. When used separately by site and treatment group, the only variable that pretty consistently predicted total unsecured debt was bill troubles.

More specifically, the procedure assumes that the log of unsecured debt (among those with any such debt) is normally distributed, with a mean that is a function of known factors and fixed variance as follows:

$$\ln Y_i \sim N\left(X_i \beta, \sigma^2\right)$$

Having estimated β and σ^2 and knowing that $Y_i \in \left[\ell, u\right]$, the exact debt amount is imputed as

$$\hat{Y}_i = F_i^{-1}\left(F_i(\ell) + p\left(F_i(u) - F_i(\ell)\right)\right)$$

where *p* is a random number between 0 and 1, *F* is the distribution function for a log-normal random variable with log-mean $X_i \hat{\beta}$ and log-variance $\hat{\sigma}^2$:

$$F_i(y) = \begin{cases} \dfrac{1}{\sqrt{2\pi\hat{\sigma}^2}} \displaystyle\int_{-\infty}^{\ln(y)} \exp\left(-\frac{\left(\nu - X_i\hat{\beta}\right)^2}{2\hat{\sigma}^2}\right) d\nu & y > 0 \\ 0 & y \le 0 \end{cases}$$

The highest category on the questionnaire was $50,000 or more. We set the upper limit to this category as $1 million. This resulted in some rare very high levels of imputed total unsecured debt. The highest was close to $450,000.

B.4 Survey Nonresponse Analysis

It is possible that the impact of a PACE program on people who respond to the follow-up survey is different than its impact on those who do not respond. This section documents analyses related to this topic as well as adjustments to inferential procedures to reduce this threat. Section B.4.1 presents some general considerations on the topic, including how both regression adjustment and weights can be useful. Section B.4.2 assesses our data for evidence of systematic nonresponse bias. We did find such evidence at one site. Section B.4.3 documents our approach to developing and testing weights to reduce potential nonresponse bias. This approach uses current data from administrative sources in addition to baseline data collected by the evaluation.

Because populations and survey field periods differed across the four sites, we analyzed and developed weights for each site separately. We tested a number of weighting procedures. Differences in the procedures' performance across sites were fairly small, so for simplicity we ended up selecting the same approach in all four sites.

B.4.1 General Considerations about Nonresponse in Follow-Up Surveys of Participants in Randomized Studies

As a rough proxy for vulnerability of inferences to nonresponse bias, some researchers focus on the overall response rate and the difference in response rates between the treatment and control groups.[56] Accordingly, we show these rates in Exhibit B-9. However, we note that regression adjustment provides strong protection against nonresponse bias on estimated impacts even when response rates vary across the treatment and control groups and even when unadjusted outcome means for the treatment and control groups suffer nonresponse bias. This is true because the regression adjustment can remove noise and bias caused by imbalance in the respondent sample, just as it removes noise on the full sample and as it removes bias in quasi-experimental designs. Regression-adjusted impact estimates are biased by nonresponse only if there are uncontrolled factors (i.e., factors beyond the set of covariates used in the regression) that cause deviations in both outcomes and nonresponse *propensity* (the probability of nonresponse). Therefore, focusing on differential response rates across the two study groups can lead to undue concern about nonresponse bias.

Instead, we focus on whether the relevant factors have been included in the regression adjustment. Of course, if there are post-randomization factors that directly influence both outcomes and nonresponse propensity, then regression adjustment cannot remove the bias because it is not permissible to include post-randomization factors in a regression adjustment. If evidence of the importance of such factors is found, then it may be useful to create nonresponse-adjustment weights and to use these weights in weighted regression analysis. (The advantage with weights is that they—unlike regression adjustment—are allowed to incorporate post-randomization information.)

Exhibit B-9: Response Rates to the Six-Year Survey, by Site

Site	Treatment Group (%)	Control Group (%)	Total (%)	T-C Difference (percentage points)	Survey-Eligible Sample Size (#)
Carreras	72.4	62.5	67.5	9.9	794
I-BEST	61.2	55.4	58.3	5.8	614
VIDA	81.5	73.1	77.3	8.4	947
Year Up	67.7	61.7	65.7	6.0	2,517

Source: PACE six-year follow-up survey.

Note: The survey-eligible sample excludes those who had died prior to the survey attempt, as well as those who were incarcerated or suffering from a serious health condition that made response impossible.

56 See for example, Deke and Chiang (2017). For a slightly contrarian view, see Hendra and Hill (2018).

B.4.2 Evidence of Nonresponse Bias in Unadjusted Impact Estimates

We used administrative data to look for evidence of nonresponse bias in both group means and impacts. We knew who responded to the survey and we had administrative data outcomes for both survey respondents and nonrespondents. We can thus compute two estimates (for either group means or impacts) from the administrative data: one estimate from the survey-eligible sample, which we treat as truth; and a second estimate from the sample of people who responded to the six-year survey. In the absence of nonresponse bias (and with large enough samples), we should get (up to sampling variability) the same estimates of group means and impacts on the full sample and on the sample of survey respondents.

As discussed in the prior section, even when nonresponse leads to bias in the estimated mean outcomes in the treatment or control group, this need not translate into bias on the estimated impact. Nonetheless, during the initial work investigating the possibility of bias, we focused on absolute bias on means for the two groups because this allowed us to minimize the potential for bias without exposing ourselves to estimated impacts and thereby perhaps tainting our impartiality.[57] In that early-phase research, using NSC data we found substantial evidence of nonresponse bias in group means. Accordingly, we developed an initial set of nonresponse-adjustment weights, as described next in Section B.4.3. This approach was very successful in removing bias from estimated group means for educational progress outcomes. Given a lack of bias, we would also expect it to produce nearly unbiased impacts on educational progress outcomes.

Nonetheless, bias in group means is far less troublesome than bias in estimated impacts. Moreover, even if weights can bring the means of the respondent sample by treatment/control status into perfect agreement with the corresponding means of the full sample, this does not automatically mean that using the weights in a regression analysis of the impact of treatment will also bring the regression-adjusted impact on the respondent sample into perfect agreement with the regression-adjusted impact on the full sample. For this reason, after preparing a preliminary set of nonresponse weights that worked well for group means of educational progress outcomes, we took a careful look at the impact of weights on nonresponse bias in regression-adjusted impact estimates using NDNH data.

It is complex to determine how much sampling variability to expect in differences between impacts estimated on the respondent sample and on the full sample. The box below indicates how much sampling variability can be expected in a pair of nested means when the difference between them is standardized by the standard error on the full-sample estimate. We use this to approximate the amount of variability that can be expected in the standardized difference of a pair of nested regression-adjusted impact estimates.[58]

---

57 By ignoring the sign of the bias in both the treatment and control groups, we had no way of telling whether the bias would lead to larger or smaller program impacts. That is the advantage of looking at the *absolute* bias in each arm and abstaining from calculation of biases on impacts.

58 We could have directly estimated this with resampling variance estimation techniques such as the bootstrap or the jackknife, but the benefit would be small relative to the time and costs.

Variance of a Nested Difference in Means

If $\bar{y}_s$ is the mean of some variable on a subset *s* of the full sample, $\bar{y}_F$ is the corresponding mean of the full sample, and the subset comprises 100*r* percent of the full sample, then

$$\begin{aligned}
\operatorname{cov}(\bar{y}_s, \bar{y}_F) &= \operatorname{cov}\left(\frac{1}{rn}\sum_{i\in s} y_i, \frac{1}{n}\sum_{i\in F} y_i\right) \\
&= \operatorname{cov}\left(\frac{1}{rn}\sum_{i\in s} y_i, \frac{1}{n}\sum_{i\in s} y_i\right) \\
&= \frac{1}{rn^2}\sum_{i\in s}\operatorname{cov}(y_i, y_i) \\
&= \frac{\sigma^2}{n}
\end{aligned}$$

So

$$\begin{aligned}
\operatorname{var}(\bar{y}_s - \bar{y}_F) &= \operatorname{var}(\bar{y}_s) + \operatorname{var}(\bar{y}_F) - 2\operatorname{cov}(\bar{y}_s, \bar{y}_F) \\
&= \frac{\sigma^2}{rn} + \frac{\sigma^2}{n} - \frac{2\sigma^2}{n} \\
&= \frac{\sigma^2}{n}\left(\frac{1}{r} + 1 - 2\right) \\
&= \frac{\sigma^2}{n}\left(\frac{1-r}{r}\right) = \left(\frac{1-r}{r}\right)\operatorname{var}(\bar{y}_F)
\end{aligned}$$

Therefore, a 95 percent confidence internal on $\mu_S - \mu_F$ (the true difference between the subsample and the full sample) extends from $\bar{y}_s - \bar{y}_F - 1.96\sqrt{\left(\frac{1-r}{r}\right)\operatorname{var}(\bar{y}_F)}$ to $\bar{y}_s - \bar{y}_F + 1.96\sqrt{\left(\frac{1-r}{r}\right)\operatorname{var}(\bar{y}_F)}$. Even with zero nonresponse bias, we would expect some difference between the two observed means due to sampling error alone. Using our confidence interval above, we can conclude that if there were no true difference in the two means, we would expect that 95 percent of the time $\frac{\bar{y}_s - \bar{y}_F}{\sqrt{\operatorname{var}(\bar{y}_F)}}$ would fall between $-1.96\sqrt{\left(\frac{1-r}{r}\right)}$ and $+1.96\sqrt{\left(\frac{1-r}{r}\right)}$. So, we have strong evidence of nonresponse bias only if $\frac{|\bar{y}_s - \bar{y}_F|}{\sqrt{\operatorname{var}(\bar{y}_F)}} > 1.96\sqrt{\frac{1-r}{r}}$.

Given the response rates at Carreras en Salud and Year Up of about 67 percent each, the equations in the box imply that there is reason to be concerned about nonresponse bias only if the standardized difference is larger than 139 percent at these two sites. At I-BEST, the response rate is lower; at that site we are concerned about nonresponse bias only if the standardized difference is larger than 163 percent. At VIDA, the response rate is higher; at that site we are concerned about nonresponse bias if the standardized difference is larger than 107 percent.

Although differences smaller than these thresholds are difficult to interpret, subsequent exhibits in this appendix also report the unstandardized impact estimates, including the sign of the impact (positive or negative) and the significance of the estimated impact. Given that the survey respondent sample size is smaller than the randomized sample size, some loss of statistical significance is expected for estimates based on just the survey respondent sample.

For each of the four sites with six-year follow-up survey data, we looked for evidence of nonresponse bias on three NDNH outcomes and three NSC outcomes. The results of this search are shown in Exhibits B-10 through B-13. Looking across the four exhibits, we see some potentially troubling differences in estimates of the impacts of Carreras en Salud, I-BEST, and VIDA, but none of the standardized differences reach the threshold for concern at the site. We see strong evidence of nonresponse bias only at Year Up.

In particular, in Exhibit B-13, we see that the impact of Year Up on cumulative FTE months of college enrollment through Q24 is 1.07 months on the full survey-eligible sample, but only 0.33 months on the survey respondent sample. The difference in impacts is 0.70 months, or 153 percent of the full sample standard error of 0.48 months. (Recall that the threshold for concern is 139 percent of full sample standard error.) This difference in impacts would be consistent with a pattern in which either highly persistent students in the treatment group were less likely to respond to the survey, or highly persistent students in the control group were more likely to respond.

The difference in estimates of the impact of Year Up on Q23/24 earnings is also close to the threshold for concern. The impact is $1,933 on the full survey-eligible sample, but $2,303 on the survey respondent sample. The difference in impacts is $370, or 137 percent of the full sample standard error of $270. This difference in impacts would be consistent with a pattern whereby either higher earners in the treatment group were more likely to respond to the survey, or higher earners in the control group were less likely to respond.

Given these findings, we tested various weighting schemes to reduce the bias, particularly at Year Up.

Exhibit B-10: Evidence of Nonresponse Bias in Estimated Impact of Carreras en Salud on Administratively Measured Outcomes

	Survey-Eligible Sample (Regression-Adjusted)				Survey Respondent Sample (Regression-Adjusted)	
Outcome	Treatment Group	Control Group	Impact	Standard Error	Impact	Standardized Absolute Difference (%)
NDNH						
Quarterly earnings (average of Q23 and Q24) ($)	6,667	6,317	350	375	394	12
Employed during Q23 (%)	78.96	74.90	4.06*	3.06	1.17	94
Cumulative earnings through Year 6 ($)	113,340	117,520	−4,180	4,679	−5,641	31
Average						46
Sample size (treatment + control group)		771			520	
NSC						
Any long-term credentials by Q24 (%)	16.94	13.16	3.78	2.41	5.06	53
Received associate or higher degree by Q24 (%)	10.79	8.86	1.93	2.02	3.49	78
Cumulative FTE months of college enrollment through Q24	6.50	5.36	1.14*	0.69	1.29	22
Average						51
Sample size (treatment + control group)		794			536	

Key: FTE = full-time-equivalent.

Source: National Directory of New Hires (data received as of March 16, 2021); National Student Clearinghouse.

Note: All estimates are regression-adjusted as discussed in Appendix Section A.2. The survey-eligible sample excludes those who had died as of the survey attempt, as well as those who were incarcerated or suffering from a serious health condition that made response impossible.

Statistical significance levels for exploratory outcomes are based on two-tailed tests. For confirmatory and secondary outcomes, statistical significance levels are based on one-tailed *t*-tests tests of positive differences between research groups for positive outcomes and negative differences for negative outcomes. Statistical significance levels are summarized as follows: *** 1 percent level; ** 5 percent level; * 10 percent level.

Exhibit B-11: Evidence of Nonresponse Bias in Estimated Impact of I-BEST on Administratively Measured Outcomes

	Survey-Eligible Sample (Regression-Adjusted)				Survey Respondent Sample (Regression-Adjusted)	
Outcome	Treatment Group	Control Group	Impact	Standard Error	Impact	Standardized Absolute Difference (%)
NDNH						
Quarterly earnings (average of Q23 and Q24) ($)	5,544	5,593	−49	453	−164	25
Employed during Q23 (%)	66.93	65.45	1.48	3.87	3.24	45
Cumulative earnings through Year 6 ($)	92,460	91,011	1,449	5,857	2,598	20
Average						30
Sample size (treatment + control group)	594				352	
NSC						
Any long-term credentials by Q24 (%)	13.47	12.46	1.01	2.70	2.82	67
Received associate or higher degree by Q24 (%)	10.88	7.21	3.66	2.30	4.47	35
Cumulative FTE months of college enrollment through Q24	8.84	6.41	2.42***	0.81	2.12	38
Average						47
Sample size (treatment + control group)	614				358	

Key: FTE = full-time-equivalent.

Source: National Directory of New Hires (data received as of March 16, 2021); National Student Clearinghouse.

Note: All estimates are regression-adjusted as discussed in Appendix Section A.2. The survey-eligible sample excludes those who had died as of the survey attempt, as well as those who were incarcerated or suffering from a serious health condition that made response impossible.

Statistical significance levels for exploratory outcomes are based on two-tailed tests. For confirmatory and secondary outcomes, statistical significance levels are based on one-tailed *t*-tests tests of positive differences between research groups for positive outcomes and negative differences for negative outcomes. Statistical significance levels are summarized as follows: *** 1 percent level; ** 5 percent level; * 10 percent level.

Exhibit B-12: Evidence of Nonresponse Bias in Estimated Impact of VIDA on Administratively Measured Outcomes

Outcome	Survey-Eligible Sample (Regression-Adjusted)				Survey Respondent Sample (Regression-Adjusted)	
	Treatment Group	Control Group	Impact	Standard Error	Impact	Standardized Absolute Difference (%)
NDNH						
Quarterly earnings (average of Q23 and Q24) ($)	8,341	8,350	−9	428	424	101
Employed during Q23 (%)	79.05	78.20	0.85	2.61	3.55	103
Cumulative earnings through Year 6 ($)	120,325	126,076	−5,751	5,792	396	106
Average						104
Sample size (treatment + control group)	944				729	
NSC						
Any long-term credentials by Q24 (%)	67.22	54.87	12.34***	2.93	11.31***	35
Received associate or higher degree by Q24 (%)	49.55	40.89	8.66***	3.07	8.40**	9
Cumulative FTE months of college enrollment through Q24	18.45	15.65	2.80***	0.80	2.28**	65
Average						36
Sample size (treatment + control group)	947				732	

Key: FTE = full-time-equivalent.

Source: National Directory of New Hires (data received as of March 16, 2021); National Student Clearinghouse.

Note: All estimates are regression-adjusted as discussed in Appendix Section A.2. The survey-eligible sample excludes those who had died as of the survey attempt, as well as those who were incarcerated or suffering from a serious health condition that made response impossible.

Statistical significance levels for exploratory outcomes are based on two-tailed tests. For confirmatory and secondary outcomes, statistical significance levels are based on one-tailed *t*-tests tests of positive differences between research groups for positive outcomes and negative differences for negative outcomes. Statistical significance levels are summarized as follows: *** 1 percent level; ** 5 percent level; * 10 percent level.

Exhibit B-13: Evidence of Nonresponse Bias in Estimated Impact of Year Up on Administratively Measured Outcomes

Outcome	Survey-Eligible Sample (Regression-Adjusted) Treatment Group	Control Group	Impact	Standard Error	Survey Respondent Sample (Regression-Adjusted) Impact	Standardized Absolute Difference (%)
NDNH						
Quarterly earnings (average of Q23 and Q24) ($)	8,787	6,854	1,933***	270	2,303***	137
Employed during Q23 (%)	82.57	82.32	0.25	1.62	0.83	36
Cumulative earnings through Year 6 ($)	138,997	108,725	30,272***	3,139	35,029	152
Average						108
Sample size (treatment + control group)	2,474				1,623	
NSC						
Any long-term credentials by Q24 (%)	12.23	13.41	−1.19	1.35	−1.77	43
Received associate or higher degree by Q24 (%)	9.59	11.33	−1.74	1.24	−3.06*	107
Cumulative FTE months of college enrollment through Q24	9.47	8.40	1.07**	0.48	0.33	153
Average						101
Sample size (treatment + control group)	2,517				1,653	

Key: FTE = full-time-equivalent.

Source: National Directory of New Hires (data received as of March 16, 2021); National Student Clearinghouse.

Note: All estimates are regression-adjusted as discussed in Appendix Section A.2. The survey-eligible sample excludes those who had died as of the survey attempt, as well as those who were incarcerated or suffering from a serious health condition that made response impossible.

Statistical significance levels for exploratory outcomes are based on two-tailed tests. For confirmatory and secondary outcomes, statistical significance levels are based on one-tailed *t*-tests tests of positive differences between research groups for positive outcomes and negative differences for negative outcomes. Statistical significance levels are summarized as follows: *** 1 percent level; ** 5 percent level; * 10 percent level.

B.4.3 Construction of Nonresponse-Adjustment Weights

As potential levers to remove nonresponse bias, we have baseline information from the Basic Information Form and Self-Administered Questionnaire, as well as current data from the NSC and NDNH. Ideally, we would have used all these information sources on a single computing platform to model nonresponse propensity. However, the NSC would not allow its data to be stored in the ACF computing environment, where we could have merged with NDNH data. We therefore developed weights using a strategy similar to the "dual-system raking" used for the three-year survey.

"Raking" is the name for iterative procedures that create weights for a sample in such a manner that marginal tabulations of the sample agree exactly with pre-specified "control" totals in multiple "dimensions." For example, raking can be used to create weights that will cause tabulations by gender, tabulations by race, and tabulations by age all to agree with pre-specified totals for gender, race, and age. In this example, gender, race, and age are the dimensions.

In the context of nonresponse, if tabulations are prepared from the full sample and raking is used on the respondents, then weighted tabulations of the respondent sample will be in perfect agreement with parallel tabulations of the full sample. This exact multi-dimensional agreement is referred to as “hyperbalance.” In the context of an experiment, if this procedure is run separately for the treatment group and control group, then hyperbalance between respondents and nonrespondents means that the weighted balance between the treatment and control groups on the respondent sample should be just as good as on the full sample.

This hyperbalance by arm means that if we estimated treatment impact on just the respondent sample with these weights but without regression adjustment, the estimated program impact on each of these hyperbalanced variables would agree exactly with corresponding program impacts estimated on the full sample. The use of regression adjustment to estimate program impacts (rather than simple mean difference between arms) means that this agreement will not be exact, but agreement should still be very good for hyperbalanced variables. Theoretically, the hyperbalance should also improve agreement (between impact estimates based on the full sample and impact estimates based on just the respondent sample) for a variety of related parallel outcomes.

Key raking variables include both categorical variables (e.g., any NSC-reported enrollment) and interval-valued variables (e.g., number of months enrolled in college according to NSC records). Including these interval-valued variables seems particularly important because many educational progress outcomes are associated with the length of study.

The need to include continuous variables in the raking is challenging because traditional raking algorithms work only with categorical variables. In contrast, the generalized raking we propose and use here can handle a mix of categorical and continuous variables.[59] For categorical variables, the procedure guarantees perfect correspondence between the respondent sample and full sample by arm on the distribution of the sample across the categories of each variable; for continuous variables, the procedure induces perfect agreement on the marginal means of each of them.

The generalized raking procedure of Folsom and associates is available in the WTADJUST procedure of SUDAAN. We refer to our system as dual-system raking because it permits raking both to NDNH information and to NSC information even though the two types of data reside on two different systems. Ignoring prior weights and some complex features to constrain minimum and maximum weights, the WTADJUST procedure creates weights of the form[60]

$$w_i = \exp\left(X_i \beta\right)$$

59 Generalized raking is most fully developed by Folsom and Singh (2000), who in turn draw on work originally proposed by Folsom (1991), Deville and Särndal (1992), and Folsom and Witt (1994). Dual raking is similar to the approach of Judkins et al. (2007) that involves the use of raking to construct weights in quasi-experimental designs.

60 For full details on how SUDAAN handles pre-existing weights and constraints on the weights, see Section 24.2.2 of the SUDAAN 11 manual (Research Triangle Institute 2012).

where X_i is a row vector of known facts about both respondents and nonrespondents, and the column vector of coefficients β is chosen such that

$$\sum_i r_i w_i X_i = \sum_i X_i$$

where r_i is a binary indicator of response status (equal to 1 for respondents and 0 for nonrespondents) if the sample includes nonrespondents and is identically equal to 1 if the sample includes only respondents.

The details of the dual-system raking procedure are as follows:

(1) We used the LASSO procedure with 10-fold cross-validation (as described in Appendix Section A.2) to select predictors of survey response by site/arm among a list of baseline variables and three post-baseline NSC variables (months of FTE enrollment, credential receipt, and long-term credential receipt—all as of three years after randomization). The baseline variables eligible for selection were those discussed in Appendix Section A.1, as well as a flag for those randomized after February 2014 and the interaction of this flag with gender.[61] Additionally, when modeling Year Up survey response, we included local Year Up site as a possible predictor. This first step was run on a file including both respondents and nonrespondents. Because this procedure was run separately for each site/arm, the variables selected varied by site and arm.

(2) We used SUDAAN/WTADJUST to develop survey weights on the Abt server that induced hyperbalance by arm for the variables selected in Step 1 and the means of five post-baseline NSC variables. Two of these NSC variables were counts on months: months with any enrollment and months of FTE enrollment. The remaining three were binary flags: any enrollment, any completions (credentials), and any long-term credentials. All five of these variables were constrained to enrollment and completions within 75 months of randomization.[62] As was the case in Step 1, this step was run on a file including both respondents and nonrespondents.

61 We included this flag in response to the COVID-19 pandemic. Those randomized after February 2014 would have had their six-year follow-up interview in March 2020 and later, when the pandemic had already begun to disrupt countless facets of daily life. We believed this disruption could influence survey response likelihood, and thus wished to account for it in our response modeling. Additionally, we wanted to explore whether the interactions of COVID with our existing baseline variables were predictors of survey response; however, the large number of these variables prevented us from using all such interactions to model site/arm survey response. Therefore, we ran a preliminary LASSO with 10-fold cross-validation using the baseline variables and their COVID interactions to predict survey response on the full survey-eligible PACE sample. This procedure selected the interaction of gender and COVID cohort; consequently, we included this variable in the selection pool for the site/arm response modeling.

62 At the time the weights were prepared, this was the longest follow-up period available across all nine PACE sites.

(3) We merged the weights from Step 2 with baseline data and follow-up survey data on the Abt server. We then passed these merged data through to a secure ACF server, where third-party ACF contractors merged our data with NDNH earnings data, removing personal identifiers from the merged dataset. We had verified that this set of NSC-adjusted weights provides nearly unbiased impact estimates for survey-based education outcomes; but after merging the weights with NDNH data, we discovered that these NSC-adjusted weights did not remove bias in survey-based impact estimates for earnings outcomes at Year Up.

(4) To remedy this, we used SUDAAN/WTADUST on the ACF server to rake the weights from Step 2 in such a manner as to attain hyperbalance by arm on variety of baseline variables (those locally predictive of NDNH earnings during quarters 23 and 24 using a cross-validated LASSO as previously discussed) as well as NDNH earnings variables (both continuous and categorical versions of earnings during quarter 23 and 24; both continuous and categorical versions of cumulative earnings through Year 6; and continuous earnings during each of Years 4, 5, and 6). We verified that these weights resulted in the standardized impact differences for the three focal outcomes below the trouble-indicated thresholds discussed above in Section B.4.1 at all four sites with survey data. This step was run on a file including both respondents and nonrespondents, excluding those ineligible for the survey as well as those with bad Social Security numbers (SSNs).

(5) We used the weights from Step 4 on the ACF server to estimate (by arm) the distributions of survey-reported earnings and binary flags for trouble making ends meet, having health insurance, being a renter, and being part of a household where someone receives means-test public benefits. Specifically, we split current earnings at $0, $6,000, $9,000, and $12,000. (We could not use finer breaks because of sample size limitations.)

(6) We again used SUDAAN/WTADUST on the ACF server to rake the weights from Step 2, but for this step we used the control totals from Step 5 rather than the NDNH totals used in Step 4. This sixth step was run on just the respondent sample. We then verified that these weights resulted in acceptable standardized differences in impacts. (These weights did not perform as well as the weights from Step 4 in reducing nonresponse bias on the respondent sample, but they still were better than the weights developed in Step 2 for Year Up.)

(7) We exported the estimated totals from Step 5 for each arm from the ACF server to the Abt server. (The data use agreement permitted the transfer of tabulations; only the export of microdata was prohibited.)

(8) We again used SUDAAN/WTADUST to rake the weights from Step 2 to the control totals from Step 5, but this time we did the raking on the Abt server rather than on the ACF server. This eighth step was run on just the respondent sample. We then merged these with NSC data on the Abt server and verified that these weights did not cause any substantial deterioration in the quality of weights for outcomes in the educational

progress domain. These weights are publishable in study archives so that our work can be replicated by others.

Exhibits B-14 through B-17 shows the performance of four sets of weights on the same critical NDNH and NSC outcomes as were the focus of Section B.4.2, where performance is judged by the absolute difference between regression-adjusted impacts on the respondent sample and corresponding regression-adjusted impacts on the full (eligible) sample. Each of the first four of these focuses on one of the four sites with a follow-up survey at six years. All impacts were estimated with regression adjustment, as discussed in Appendix Section A.2. The weights highlighted are these: no weights, the initial set of weights developed without NDNH information in Step 2 above, the set of weights developed with full access to NDNH data in Step 4 above, and the final set of weights in Step 8 that are publishable.

Examining Exhibit B-14, we see that the initial set of weights for the Carreras sample based on baseline project data and current NSC data modestly reduce the risk of nonresponse bias for outcomes in both the labor market domain and the educational progress domain. For example, the standardized absolute difference in impact estimates for Q23/Q24 quarterly earnings is reduced from 12 percent to 0 percent; meaning that with these weights, the estimated impact of Carreras on current quarterly earnings is nearly exactly the same on the respondent sample as on the survey-eligible sample. Similarly, the standardized absolute difference in impact estimates on earning a long-term credential by Q24 is reduced from 53 percent to 7 percent. The inclusion of current earnings data into nonresponse models did not materially reduce the standardized absolute impacts in earnings any further. Nonetheless, as mentioned previously, the weights from Step 8 were used as final weights to prepare estimates of the impact of Carreras on survey-measured outcomes.

Exhibit B-14: Performance of Alternate Weights for Carreras en Salud

Outcome	Standardized Absolute Difference between the Impact Estimated on Survey Respondents and the Full Survey-Eligible Sample: Unweighted	Without NDNH (Step 2)	With NDNH on ACF Platform (Step 4)	Publishable (Step 8)
NDNH				
Quarterly earnings (average of Q23 and Q24) ($)	12	0	17	12
Employed during Q23 (%)	94	79	66	83
Cumulative earnings through Year 6 ($)	31	17	24	2
Average	46	32	36	32
Sample sizes (treatment + control group): Survey-Eligible Sample 771 Survey Respondents 520				
NSC				
Any long-term credentials by Q24	53	7	na	0
Received AA or higher degree by Q24 (%)	77	33	na	37
Cumulative FTE months of college enrollment through Q24	22	17	na	13
Average	50	19	na	17
Sample sizes (treatment + control group): Survey-Eligible Sample 794 Survey Respondents 536				

Key: FTE = full-time-equivalent. na = not applicable.
Source: National Directory of New Hires (data received as of March 16, 2021); National Student Clearinghouse.
Note: All impact estimates in this table are regression-adjusted. The survey-eligible sample excludes those who had died as of the survey attempt, as well as those who were incarcerated or suffering from a serious health condition that made response impossible. The standardized absolute difference is calculated by dividing the absolute difference in impacts by the standard error on the estimated impact of the program on the full survey-eligible sample. Step 2 weights are based on baseline project data and current NSC data. Step 4 weights are based on baseline project data and current NDNH data. Step 8 weights are based on baseline project data, current NSC data, and the adjusted distribution of survey-reported earnings from Step 4 weights.

Examining Exhibit B-15, we see that the initial set of weights for the I-BEST sample based on baseline project data and current NSC data do not materially alter the risk of nonresponse bias for outcomes in either the labor market domain or the educational progress domain. For example, the standardized absolute difference in impact estimates for Q23/Q24 quarterly earnings actually increases from 25 percent to 77 percent. Somewhat more encouragingly, the standardized absolute difference in impact estimates on earning a long-term credential by Q24 is reduced from 66 percent to 39 percent. The inclusion of current earnings data into nonresponse models did not materially reduce the standardized absolute impacts in earnings any further. Nonetheless, as mentioned previously, the weights from Step 8 were used as final weights to prepare estimates of the impact of I-BEST on survey-measured outcomes.

Exhibit B-15: Performance of Alternate Weights for I-BEST

	Standardized Absolute Difference between the Impact Estimated on Survey Respondents and the Full Survey-Eligible Sample			
Outcome	**Unweighted**	**Without NDNH (Step 2)**	**With NDNH on ACF Platform (Step 4)**	**Publishable (Step 8)**
NDNH				
Quarterly earnings (average of Q23 and Q24) ($)	25	77	4	86
Employed during Q23 (%)	45	4	29	36
Cumulative earnings through Year 6 ($)	20	4	14	9
Average	30	29	16	44
Sample sizes (treatment + control group): Survey-Eligible Sample 594 Survey Respondents 352				
NSC				
Any long-term credentials by Q24	66	39	na	28
Received AA or higher degree by Q24 (%)	35	47	na	40
Cumulative FTE months of college enrollment through Q24	37	53	na	61
Average	46	46	na	43
Sample sizes (treatment + control group): Survey-Eligible Sample 614 Survey Respondents 358				

Key: FTE = full-time-equivalent. na = not applicable.
Source: National Directory of New Hires (data received as of March 16, 2021); National Student Clearinghouse.
Note: All impact estimates in this table are regression-adjusted. The survey-eligible sample excludes those who had died as of the survey attempt, as well as those who were incarcerated or suffering from a serious health condition that made response impossible. The standardized absolute difference is calculated by dividing the absolute difference in impacts by the standard error on the estimated impact of the program on the full survey-eligible sample. Step 2 weights are based on baseline project data and current NSC data. Step 4 weights are based on baseline project data and current NDNH data. Step 8 weights are based on baseline project data, current NSC data, and the adjusted distribution of survey-reported earnings from Step 4 weights.

Examining Exhibit B-16 below, we see that the initial set of weights for the VIDA sample based on baseline project data and current NSC data modestly reduce the risk of nonresponse bias for outcomes in both the labor market domain and the educational progress domain. For example, the standardized absolute difference in impact estimates for Q23/Q24 quarterly earnings is reduced from 101 percent to 92 percent, still a number that is close to threshold for statistical significance. More encouragingly, the standardized absolute difference in impact estimates on earning a long-term credential by Q24 is reduced from 35 percent to 13 percent. The inclusion of current earnings data into nonresponse models further reduces the standardized absolute impacts in earnings while slightly increasing it for outcomes in the educational progress domain. Given the importance of labor market outcomes at six years, we decided to use the Step 8 weights.

Exhibit B-16: Performance of Alternate Weights for VIDA

	Standardized Absolute Difference between the Impact Estimated on Survey Respondents and the Full Survey-Eligible Sample			
Outcome	Unweighted	Without NDNH (Step 2)	With NDNH on ACF Platform (Step 4)	Publishable (Step 8)
NDNH				
Quarterly earnings (average of Q23 and Q24) ($)	101	92	11	27
Employed during Q23 (%)	103	87	23	63
Cumulative earnings through Year 6 ($)	106	90	12	31
Average	104	90	15	40
Sample sizes (treatment + control group): Survey-Eligible Sample 944 Survey Respondents 729				
NSC				
Any long-term credentials by Q24	35	13	na	38
Received AA or higher degree by Q24 (%)	8	5	na	19
Cumulative FTE months of college enrollment through Q24	65	2	na	0
Average	36	7	na	19
Sample sizes (treatment + control group): Survey-Eligible Sample 947 Survey Respondents 732				

Key: FTE = full-time-equivalent. na = not applicable.
Source: National Directory of New Hires (data received as of March 16, 2021); National Student Clearinghouse.
Note: All impact estimates in this table are regression-adjusted. The survey-eligible sample excludes those who had died as of the survey attempt as well as those who were incarcerated or suffering from a serious health condition that made response impossible. The standardized absolute difference is calculated by dividing the absolute difference in impacts by the standard error on the estimated impact of the program on the full survey-eligible sample. Step 2 weights are based on baseline project data and current NSC data. Step 4 weights are based on baseline project data and current NDNH data. Step 8 weights are based on baseline project data, current NSC data, and the adjusted distribution of survey-reported earnings from Step 4 weights.

Examining Exhibit B-17 below, we see that the initial set of weights for the Year Up sample based on baseline project data and current NSC data modestly reduce the risk of nonresponse bias for outcomes in the labor market domain while strongly reducing it in the educational progress domain. For example, the standardized absolute difference in impact estimates for Q23/Q24 quarterly earnings is reduced from 137 percent to 103 percent, still a number that is close to threshold for statistical significance. More encouragingly, the standardized absolute difference in impact estimates on earning a long-term credential by Q24 is reduced from 43 percent to 7 percent. The inclusion of current earnings data into nonresponse models further reduces the standardized absolute impacts in earnings while slightly increasing it for outcomes in the educational progress domain. Given the importance of labor market outcomes at six years, the team decided to use the Step 8 weights for Year Up.

Exhibit B-17: Performance of Alternate Weights for Year Up

Outcome	Standardized Absolute Difference between the Impact Estimated on Survey Respondents and the Full Survey-Eligible Sample Unweighted	Without NDNH (Step 2)	With NDNH on ACF Platform (Step 4)	Publishable (Step 8)
NDNH				
Quarterly earnings (average of Q23 and Q24) ($)	137	103	3	39
Employed during Q23 (%)	36	9	20	15
Cumulative earnings through Year 6 ($)	152	89	1	36
Average	108	67	8	30
Sample sizes (treatment + control group): Survey-Eligible Sample 2,474 Survey Respondents 1,623				
NSC				
Any long-term credentials by Q24	43	7	na	14
Received AA or higher degree by Q24 (%)	106	35	na	40
Cumulative FTE months of college enrollment through Q24	153	14	na	16
Average	101	19	na	23
Sample sizes (treatment + control group): Survey-Eligible Sample 2,517 Survey Respondents 1,653				

Key: FTE = full-time-equivalent. na = not applicable.
Source: National Directory of New Hires (data received as of March 16, 2021); National Student Clearinghouse.
Note: All impact estimates in this table are regression-adjusted. The survey-eligible sample excludes those who had died as of the survey attempt as well as those who were incarcerated or suffering from a serious health condition that made response impossible. The standardized absolute difference is calculated by dividing the absolute difference in impacts by the standard error on the estimated impact of the program on the full survey-eligible sample. Step 2 weights are based on baseline project data and current NSC data. Step 4 weights are based on baseline project data and current NDNH data. Step 8 weights are based on baseline project data, current NSC data, and the adjusted distribution of survey-reported earnings from Step 4 weights.

Appendix C: National Student Clearinghouse Data

The National Student Clearinghouse (NSC) is a national database of college enrollment records designed to aid the administration of student loan programs. The NSC is also a useful tool for education researchers. In addition to using it to measure key outcomes for the impact analysis, this report used NSC records in some imputations for missing data. Section C.1 summarizes statistics on NSC coverage. Section C.2 provides details on how raw data from NSC were recoded for outcomes used in the impact analysis.

C.1 Coverage

Given its focus on loan administration, NSC covers only Title IV schools; that is, the set of schools approved for federal student loans by the U.S. Department of Education. Moreover, although NSC does include a few schools that are not "colleges" in the sense used elsewhere in this report (i.e., issuing degrees), the vast majority of the schools are colleges. Exhibit C-1 shows the percentage of colleges providing records to the NSC by year and by type of school. As shown, coverage of public two-year and four-year schools was more than 95 percent. Coverage was lower among private not-for-profit four-year schools, considerably lower among private for-profit four-year schools, and very low for private two-year schools (both for-profit and not-for-profit).

Exhibit C-1: NSC College-Level Cooperation Rates by College Control and Level, 2013-2016

Type and Control of College	2013 (%)	2014 (%)	2015 (%)	2016 (%)
Public, four-year	99.2	99.4	99.5	99.6
Private not-for-profit, four-year	93.6	95.2	95.8	96.1
Private for-profit, four-year	74.4	79.9	81.7	81.0
Public, two-year	99.1	99.2	99.4	99.5
Private not-for-profit, two-year	39.5	40.8	40.4	42.1
Private for-profit, two-year	19.7	28.1	26.7	26.6

Source: National Student Clearinghouse https://nscresearchcenter.org/wp-content/uploads/NSC_COVERAGE.xlsx.

Analyses of NSC data in this report are limited to enrollment records from 2000 forward.[63] All study participants gave their informed consent to have NSC share their records with the PACE study. The research team negotiated a contract with the NSC to match relevant NSC records to the study participants. The team sent both SSNs and names to NSC to make the matching more accurate. NSC then sent the abstracted records by encrypted secure methods back to the research team, which has used them under tight security conditions.

63 NSC has older records, but we did not request them as we did not need those data.

C.2 Data and Measures

Counting the quarter during which random assignment occurred as Q0, for all PACE participants, the NSC provided an extract in March 2021 covering enrollment and credentials through at least Q24 for all study participants. Exhibit C-2 shows the number of NSC quarters for which we have complete data for each site as of this writing. This number varies across sites because of the variation in the end dates for study enrollment and randomization across the sites. We obtained a new NSC extract for VIDA and Year Up in August 2021 in order to stretch the number of quarters with complete data to 27 for those impact reports. But we used the March 2021 NSC extract for all analysis in this appendix volume.

Although this report makes use of both enrollment and credential data, NSC documentation and other research indicate that the credential data are less complete than the enrollment data.[64]

Records from the NSC are arranged in a spell format with starting and ending dates. We translated these first into a set of person-month-level records, reconciling multiple and conflicting spells as seemed most sensible. We derived three variables for each person-month. The first was a simple binary indicator of “any enrollment.” The second was a binary indicator of “any full-time enrollment.” The third was a measure of full-time-equivalent (FTE) enrollment that took the values 1 (for full-time enrollment), 0.75 for three-quarter-time enrollment, 0.5 for half-time enrollment, 0.25 for some but less than half-time enrollment, and 0 for no enrollment.[65]

Translating these to person-quarter-level outcomes, a student counted as enrolled for the quarter if they were enrolled in any of the three months of that quarter, enrolled full-time if they were enrolled full-time in any of the three months of that quarter, and FTE enrollment was calculated by summing the student’s total FTE months for the quarter.

---

[64] Dundar and Shapiro (2016) indicate that schools that choose to submit information on type of credential pursued or earned do so voluntarily and with minimal processing by NSC staff. About 90 percent of students attend schools that do submit information on credential types, but there is no systematic classification scheme for credentials that are not degrees. Schools merely submit names of certificates and diplomas awarded. The authors also specifically note that information on earned credits is weak. In addition, Dynarski, Hemelt, and Hyman (2015) report that only about 80 percent of degrees from Michigan colleges were reported to the NSC in the 2008-2010 period.

[65] Because informed consent had been collected from all study participants, the NSC shared full/part-time status for everyone in the sample, something that is not otherwise shared with researchers. The factors of 0.75, 0.50, and 0.25 were our own translation.

Exhibit C-2: Length of Follow-Up with NSC Data

Site	Number of Complete Quarters Available for Appendix Volume	Planned Number of Complete Quarters
Bridge to Employment in the Healthcare Industry (BTH)	27	27
Carreras en Salud (Carreras)	25	25
Health Careers for All (HCA)	24	24
Integrated Basic Education and Skills Training (I-BEST)	25	25
Patient Care Pathway Program (PCPP)	27	27
Pathways to Healthcare (PTH)	27	27
Workforce Training Academy Connect (WTAC)	24	24
Valley Initiative for Development and Advancement (VIDA)	26	27
Year Up (Year Up)	25	27

In addition, we specified the following outcomes to capture substantial and/or sustained educational progress:

- Cumulative months with any enrollment.[66]
- Cumulative months of FTE enrollment.
- Cumulative months with any full-time enrollment.[67]
- Earned a degree.[68]
- Earned a degree or earned a credential preceded by eight or more FTE months of enrollment.[69]
- Earned a credential and subsequently enrolled for four or more months.
- Earned a degree or earned a credential preceded by eight or more FTE months of enrollment and subsequently enrolled for four or more months.
- Any college enrollment after Year 3.[70]
- Earned any credential after Year 3.

As discussed in Section B.2, we also used NSC data to help impute the required length of study for some college-issued certificates.

66 Cumulative months with any college enrollment through Year 7 is a secondary outcome for VIDA.

67 Cumulative months with any full-time college enrollment through Year 7 is a secondary outcome for VIDA.

68 A secondary outcome for PCPP, PTH, and VIDA.

69 A confirmatory outcome at six years for Carreras, I-BEST, PCPP, and VIDA.

70 A secondary outcome for PTH.

Reports for several of the programs place particular emphasis on earning a degree or other credential after eight or more FTE months of college enrollment. This is a proxy for the originally declared confirmatory outcome of earning a survey-reported credential requiring at least a full year or more's worth of credit. As discussed in Section B.2 above, the survey-based classification was problematic for credentials for which the only available information was the name of the credential. A similar problem was encountered in the NSC because most schools report only credential names and award dates to the NSC, with no information on the required length of study.

However, we determined that combining FTE months of enrollment with award dates allowed us to create a reasonable proxy. We judged this by comparing this measure at 35 months with the college records at VIDA, a site where almost all college study was at colleges that had shared their records with the study.[71] These college records allowed a very accurate classification of the length of required study. As shown in Exhibit C-3, agreement between the NSC proxy and the best estimate based on local college records is very good. The marginal rates are very similar (87 percent according to the NSC proxy versus 89 percent in the college records) and the two disagree for only 14 percent of the sample.

Exhibit C-3: Agreement of NSC-Based Proxy for a Longer-Term Certificate, Diploma, or Degree versus Local College Records, Study Participants at VIDA

	NSC Proxy		
Local College Records	**Study Member Earned a Degree or Credential After at Least 8 FTE Months of Enrollment**	**Study Member Did not Earn Such Degree or Credential**	**Total**
Study Member Earned a Degree or Credential Requiring at Least a Year of Study	435	56	**491**
Study Member Did not Earn Such Degree or Credential	67	305	**372**
Total	**502**	**361**	**863**

Source: National Student Clearinghouse; VIDA partner college records.

We also considered the possibility of a proxy based on 12 or more months of study with no adjustment for part-time enrollment, but that proxy (not shown) did not agree as well with the local records as did the NSC proxy. Also not shown, we compared the NSC proxy to the estimates for the earlier reports on impacts at three years. Agreement was good at all sites except the Pathways to Healthcare program at Pima Community College. As reported in the appendix to its report (Judkins, Litwok, and Gardiner 2020, Section D.3), there seems to have been some inconsistencies in the transmission of Pima records to the NSC during the first two years of the operation of the program.

71 More than 99 percent of the sample with any college enrollment had at least one spell of enrollment at one of the VIDA-partner colleges. Nonetheless, the local college records are expected to slightly under-report relative to the NSC because 4 percent of study participants with any college enrollment had a spell at both a VIDA-partner college and some other college. (Judkins et al. 2021, Section B.1)

Appendix D: Unemployment Insurance Wage Detail

Through the 1990s, many social program evaluations relied on administrative earnings data provided by state Unemployment Insurance (UI) agencies. State agencies maintained these data, and privacy concerns sometimes precluded sharing them with outside researchers. UI records have become more accessible since 1996 with the advent of a centralized national database—the National Directory of New Hires (NDNH). Among the NDNH's virtues is that, unlike state data, it captures earnings for study participants who work for the federal government, work in a different state than their state of residence, or move to another state during the follow-up period.

The federal Office of Child Support Enforcement (OCSE) in the U.S. Department of Health and Human Services' Administration for Children and Families (ACF) operates the NDNH.[72] The NDNH contains new hire, quarterly wage, and UI information submitted by State Directories of New Hires, employers, and state workforce agencies. The NDNH also includes the state reports with records about earnings from federal civilian and military jobs (which are otherwise not covered by state UI data). Given this supplementation, the most important sources of uncaptured earnings are from self-employment, firms' employment of independent contractors, unreported tips, and informal employment.[73]

D.1 Data Collection Process

OCSE primarily uses NDNH to assist state child support agencies locate parents, establish paternity, and collect child support. In addition, subject to federal law, regulation, guidance, and other requirements to protect data privacy and security,[74] OCSE may disclose certain information contained in the NDNH to requesting local, state, or federal agencies for research likely to contribute to achieving the purposes of part A or part D of title IV of the Social Security Act. Part A governs the federal Temporary Assistance for Needy Families (TANF) program. Part D governs the state/federal child support program. Such disclosures may not include the names, Social Security numbers (SSNs), or other personally identifying information.

If the disclosure is approved, the agency and OCSE must work together on the operational issues surrounding the technical and procedural aspects of the disclosure, such as mitigating the risks of identifiability and establishing appropriate data retention and disposition schedules of data files.

---

72 For more information, see OCSE's Guide to NDNH document: https://www.acf.hhs.gov/sites/default/files/documents/ocse/a_guide_to_the_national_directory_of_new_hires.pdf

73 According to the U.S. Bureau of Labor Statistics, about 10 percent of workers are self-employed: https://www.bls.gov/spotlight/2016/self-employment-in-the-united-states/home.htm

74 The legal authority for this disclosure for research purposes is contained in subsection 453(j)(5) of the Social Security Act.

ACF's Office of Planning, Research, and Evaluation (OPRE) and OCSE negotiated a memorandum of understanding allowing access to NDNH data for the PACE project. Among other provisions, the memorandum dictates what self-reported data from study subjects may be merged with NDNH data, the computing environment where these merges are conducted, and procedures for review of analysis results prior to release.

The PACE research team transmits match request files to OCSE quarterly. These match request files contain the names and SSNs of PACE study participants. OCSE verifies with the Social Security Administration that the reported SSNs belong to the named persons. For those SSNs that pass this test, OCSE copies NDNH records for that quarter and the preceding seven quarters to a secure folder on the ACF server.[75] (Ordinarily, these records would be destroyed after two years.) These copied records contain a pseudo-SSN; the records are stripped of all personal identifiers.

States are required to submit earnings records to OCSE within four months, but there are stragglers and corrections. To account for possible delays, PACE analyses limit NDNH-based measures to time periods that ended at least six months prior to the extract date.

Once the study is ready to analyze the collected data, the study submits a "passthrough" file to OCSE containing a variety of PACE-assigned variables (such as treatment status and program ID) and self-reported variables (such as the baseline information described in Appendix A). OCSE then strips the personal identifiers out of the passthrough file and replaces the actual SSNs with the same pseudo-SSNs previously assigned to the archived wage records. The study then uses these pseudo-SSNs to merge program and self-reported data with NDNH quarterly wage data on ACF's secure server in order to estimate program impacts on earnings and employment.

D.2 Data and Measures

Random assignment for the nine PACE programs began and ended at different points. Thus, wage records from NDNH were available for differing numbers of post-randomization quarters for each site. Exhibit D-1 below displays the start and end dates of randomization, as well as the number of post-randomization quarters of wage data available for each of the nine PACE sites.[76] In addition to the quarters of post-randomization data detailed below, we had eight quarters of pre-randomization data for the entire sample (we included the four most recent pre-randomization quarters in our regression-adjustment models).

75 Those study participants who are not matched in the Social Security Administration database are considered "missing" for these purposes, because their employment records are not available.

76 NDNH wage records are not available for the most recent quarters prior to the publication of this report due to the lag of up to six months in processing of employer reports by states and transfer of state data to OSCE.

Exhibit D-1: Quarters of NDNH Data Used in PACE Long-Term Reports, by Site

Site	Random Assignment Start	Random Assignment End	Most Recent NDNH Records Used in Reports	# of Quarters of NDNH Data Used in Long-Term Reports: Earliest Randomized Participants	Latest Randomized Participants
Bridge to Employment in the Healthcare Industry	June 2012	October 2013	Q3 2020	33	27
Carreras en Salud	November 2011	September 2014	Q3 2020	35	24
Health Careers for All	September 2012	December 2014	Q4 2020	33	24
I-BEST	November 2011	September 2014	Q3 2020	35	24
Pathways to Healthcare	February 2012	January 2014	Q3 2020	34	26
Patient Care Pathways Program	December 2011	January 2014	Q3 2020	35	26
VIDA	November 2011	June 2014	Q1 2021	37	27
Workforce Training Academy Connect	April 2012	December 2014	Q4 2020	34	24
Year Up	January 2013	September 2014	Q2 2021	33	27

Source: National Directory of New Hires. Basic Information Form.
Note: Most numbers in this appendix cover a six-year follow-up period, so do not use all available data. The individual impact reports cover beyond six years when available and relevant to the analysis.

Of the 9,242 treatment and control group members randomized as part of this PACE six-year evaluation, 9,071 study participants reported names and SSNs that OCSE deemed to be of sufficient quality for its matching purposes.[77] Analyses in this report thus are based on the 98 percent of the sample the agency deemed suitable. Valid SSN rates varied by site and are displayed in Exhibit D-2 below.

Exhibit D-2: NDNH Sample Size Information, by Site

Site	Original Randomized Sample	Withdrawals	Six-Year Analysis Sample	Valid SSN Sample	Valid SSN Percentage (%)
Bridge to Employment in the Healthcare Industry	1,007	3	1,004	974	97
Carreras en Salud	800	1	799	775	97
Health Careers for All	654	2	652	648	>99
I-BEST	632	1	631	610	97
Pathways to Healthcare	1,220	3	1,217	1,208	>99
Patient Care Pathways Program	500	1	499	486	97
VIDA	959	1	958	955	>99
Workforce Training Academy Connect	943	0	943	920	98
Year Up	2,544	5	2,539	2,495	98
Total	**9,259**	**17**	**9,242**	**9,071**	**98**

Source: National Directory of New Hires.
Note: The count of 9,242 is smaller than the original randomized count of 9,259 because 17 participants withdrew their consent for the study to use their data in any way.

77 The acceptability of the combination of a name and an SSN can vary over time. OCSE reviews the SSN ownership every quarter for the entire sample.

This sample's earnings in each quarter were based on earnings records found for each sample member in matching. As usual in use of such data, the study defined sample members as not working when the SSN-name combination was considered valid, but there was no match to wage records in a given quarter.

Each quarter the study submitted a match request file to OCSE that contained the names and SSNs for everyone randomized to that date. Where the SSNs and names aligned, OCSE returned earnings data for the eight most recent quarters in the NDNH, which is lagged by two quarters from the date of the match. As a result, the study had up to eight wage reports for each quarter. We used the last version for each quarter within a window. For example, for earnings in the second quarter of 2014, the study used reports from the match file for the third quarter of 2016 and discarded the seven earlier sets of earnings data for the second quarter of 2014.

When the earnings data for a quarter contained two or more reports for the same person from the state, the study assumed that these reports reflected either different payments by the same employer or payments from different employers. In those situations, the two earnings reports were summed to represent the persons' quarterly earnings. Consistent with the logic discussed in Appendix F, we reviewed quarterly earnings for any values that were clearly impossible. We found one such case and reset earnings for it to missing.

We calculated two outcomes for each quarter: a binary indicator of employment (i.e., Y/N to having any earnings) and the total reported earnings for the quarter ($). The result was two series—employment and earnings—of measures for each person for: the four quarters before randomization, the quarter of randomization, and the quarters after randomization. The number of quarters available after randomization varied by site. In addition, we formed a quarterly average for Q23 and Q24 after random assignment (the confirmatory earnings outcome, established to align with the theory of change) and various annual averages.[78]

[78] The exact annual averages reported (e.g., "Years 2-6" versus "Years 4-6") differed by site.

Appendix E: Sensitivity Analyses

This appendix reports sensitivity analyses designed to explore the consequences of various decisions made throughout the PACE six-year follow-up analysis. These decisions include regression adjustment, weighting, imputation, and use of the National Student Clearinghouse (NSC) and National Directory of New Hires (NDNH) data versus data from the PACE six-year follow-up survey.[79] Throughout this appendix, we refer to changes in the numbers of "stars" attached to alternate impact estimates. We focus on stars rather than *p*-values because discussion rules used by report authors do involve stars. Obviously, all the *p*-values will change with even slight differences in estimated impacts or their associated standard errors. However, these changes in *p*-values do not change the report text unless the number of stars changes.

E.1 Regression Adjustment Sensitivity Analysis

As discussed in Appendix Section A.2, the impacts presented in the main report are regression-adjusted to improve precision and decrease the influence of sampling error on impact point estimates. Exhibits E-1a to E-1i display the effects of regression adjustment. As expected, in general, regression adjustment shrank standard errors. Notably, regression adjustment shrank the standard error of the impact estimate for the "target" outcome in each domain in each site.[80] In order to obtain variance reduction on every estimate, it would likely be necessary to run a separate least absolute shrinkage and selection operator (LASSO) for each outcome. However,

[79] All NDNH figures in the appendix volume are based on NDNH data received in either March or June 2021 (specific month indicated in table notes). This causes some—almost always small—discrepancies with numbers in the individual site reports and cross-site report. Six-year impact estimates in those reports for Bridge to Employment, Carreras, I-BEST, Pathways to Healthcare, and Patient Care Pathways Program are based on NDNH data received in March 2021. Estimates of the six-year impacts in those reports for VIDA, Health Careers for All, and Workforce Training Academy Connect are from NDNH data received in June 2021. Estimates in the six-year report for Year Up are estimated from NDNH data received in January 2022. The NDNH extract used for the reports was based on the first available data when everyone in that program's sample had at least 24 quarters of follow-up NDNH data (except for Year Up, which was based on when everyone had 27 quarters of follow-up NDNH data). The NDNH data for the appendix volume was based on most recent available data when the analyses for the appendix were conducted.

[80] As discussed in Appendix Section A.2, for a given site, we divided all outcomes into three domains (earnings and employment, educational progress, and other). For a given site, we used the LASSO with 10-fold cross-validation to select covariates for the most salient outcome in each domain. The selected covariates for that most salient outcome then became the covariates for all outcomes in that domain for the site. The most salient earnings and employment outcome for all sites was *average quarterly earnings in Q23-Q24* after randomization. The most salient educational progress outcome for Pima's Pathways to Healthcare was *receipt of a college degree by Q24*. The most salient educational progress outcome for all other sites was *receipt of a credential after 8 or more months of college enrollment by Q24*. The most salient outcome in the "other" domain was *household income in the month prior to interview*, although this is only applicable for the four surveyed sites (Carreras, I-BEST, VIDA, and Year Up).

as previously discussed, this approach would have hindered efforts toward transparency and reproducibility and was therefore not pursued.

Despite the reduction of standard errors, regression adjustment did not lead to any more significant stars on important outcomes. In fact, it reduced the number of stars on some impact estimates at Carreras and I-BEST. That these two sites lost some important stars may be due to their rather small sample sizes. Given those small sample sizes, small changes in methodology can cause point estimates to shift just past thresholds for significance stars.

E.1.1 Four Surveyed Sites

Exhibit E-1a: Comparison of Confirmatory and Secondary Impact Estimates, Unadjusted and Adjusted for Baseline Imbalances – Carreras en Salud

Domain (Data Source), Outcome	Unadjusted Estimate		Adjusted Estimate	
	Impact	Standard Error	Impact	Standard Error
Confirmatory Outcome: Earnings/Employment (NDNH)	**Full Sample**			
Average quarterly earnings Q23-Q24 after randomization ($)	+251	387	+206	368
Selected Outcomes: Earnings/Employment (NDNH)	**Full Sample**			
Employed in Q23 after randomization (%)	+4.6	3.1	+4.9	3.1
Total earnings in Years 1-6 after randomization ($)	−6,519	5,526	−5,352	4,587
Secondary Outcomes: Employment (Survey)	**Survey Respondents without Weights**			
Working full-time (35+ hours/week) (%)	+2.8	4.3	+1.2	4.3
Working at job in a program target occupation (%)	+6.6**	4.0	+5.9*	3.9
Working at job providing all of a list of five possible benefits (%)	+1.7	4.4	+0.2	4.3
Access to career network (count of 6 yes/no items)	+0.2	0.2	+0.1	0.2
Confirmatory Outcome: Educational Progress (NSC)	**Full Sample**			
Received a degree or some other credential after 8+ months of college enrollment by Q24 (Carreras, I-BEST, PCPP, VIDA) (%)	+5.1**	2.6	+3.8*	2.4
Secondary Outcomes: Financial Well-Being (Survey)	**Survey Respondents without Weights**			
Ability to handle financial emergency of $400 from savings or checking (%)	−7.4	4.2	−9.7	4.1
Average total unsecured debt ($)	+162	1,647	−147	1,646
Received means-tested public benefits (%)	−0.8	4.3	+0.7	4.3
Extent of financial distress (9-item scale, ranging from 0 to 9)	−0.1	0.1	−0.1	0.1
Selected Outcome: Financial Well-Being (Survey)	**Survey Respondents without Weights**			
Household income in month prior to interview ($)	−111	193	−154	187
Sample sizes (treatment + control group): NDNH 775 NSC 799 Survey 536				

Source: National Directory of New Hires (data received as of June 15, 2021). National Student Clearinghouse. PACE six-year follow-up survey.

Statistical significance levels for exploratory outcomes are based on two-tailed tests. For confirmatory and secondary outcomes, statistical significance levels are based on one-tailed *t*-tests tests of positive differences between research groups for positive outcomes and negative differences for negative outcomes (such as student debt). Statistical significance levels are summarized as follows: *** 1 percent level; ** 5 percent level; * 10 percent level.

Exhibit E-1b: Comparison of Confirmatory and Secondary Impact Estimates, Unadjusted and Adjusted for Baseline Imbalances – I-BEST

Domain (Data Source), Outcome	Unadjusted Estimate		Adjusted Estimate	
	Impact	Standard Error	Impact	Standard Error
Confirmatory Outcome: Earnings/Employment (NDNH)	**Full Sample**			
Average quarterly earnings Q23-Q24 after randomization ($)	+165	470	+152	443
Selected Outcomes: Earnings/Employment (NDNH)	**Full Sample**			
Employed in Q23 after randomization (%)	+3.5	3.9	+4.1	3.8
Total earnings in Years 1-6 after randomization ($)	+4,491	6,509	+4,366	5,684
Secondary Outcomes: Employment (Survey)	**Survey Respondents without Weights**			
Working full-time (35+ hours/week) (%)	+8.4*	5.3	+6.2	5.2
Working at job in a program target occupation (%)	+10.1***	3.1	+8.8***	3.2
Working at job providing all of a list of five possible benefits (%)	+7.0*	5.1	+3.6	5.1
Access to career network (count of 6 yes/no items)	+0.2	0.2	+0.1	0.2
Confirmatory Outcome: Education (NSC)	**Full Sample**			
Received a degree or some other credential after 8+ months of college enrollment by Q24 (Carreras, I-BEST, PCPP, VIDA) (%)	+2.6	2.7	+1.1	2.6
Secondary Outcomes: Financial Well-Being (Survey)	**Survey Respondents without Weights**			
Ability to handle financial emergency of $400 from savings or checking (%)	+8.1*	5.3	+7.1*	5.2
Average total unsecured debt ($)	+4,320	5,973	+3,425	6,036
Received means-tested public benefits (%)	+0.0	5.3	+2.6	5.0
Extent of financial distress (9-item scale, ranging from 0 to 9)	−0.3**	0.2	−0.2*	0.2
Selected Outcome: Financial Well-Being (Survey)	**Survey Respondents without Weights**			
Household income in month prior to interview ($)	+214	295	+30	281
Sample sizes (treatment + control group): NDNH 610 NSC 631 Survey 356				

Source: National Directory of New Hires (data received as of June 15, 2021). National Student Clearinghouse. PACE six-year follow-up survey.

Statistical significance levels for exploratory outcomes are based on two-tailed tests. For confirmatory and secondary outcomes, statistical significance levels are based on one-tailed *t*-tests tests of positive differences between research groups for positive outcomes and negative differences for negative outcomes (such as student debt). Statistical significance levels are summarized as follows: *** 1 percent level; ** 5 percent level; * 10 percent level.

Exhibit E-1c: Comparison of Confirmatory and Secondary Impact Estimates, Unadjusted and Adjusted for Baseline Imbalances – VIDA

Domain (Data Source), Outcome	Unadjusted Estimate		Adjusted Estimate	
	Impact	Standard Error	Impact	Standard Error
Confirmatory Outcome: Earnings/Employment (NDNH)	**Full Sample**			
Average quarterly earnings Q23-Q24 after randomization ($)	+400	450	+45	419
Selected Outcomes: Earnings/Employment (NDNH)	**Full Sample**			
Employed in Q23 after randomization (%)	+2.6	2.6	+1.1	2.6
Total earnings in Years 1-6 after randomization ($)	−324	6,166	−5,418	5,674
Secondary Outcomes: Employment (Survey)	**Survey Respondents without Weights**			
Working full-time (35+ hours/week) (%)	+2.8	3.5	+0.6	3.5
Working at job providing all of a list of five possible benefits (%)	+5.3	3.7	+4.1	3.8
Access to career network (count of 6 yes/no items)	+0.3**	0.2	+0.3**	0.2
Confirmatory Outcome: Education (NSC)	**Full Sample**			
Received a degree or some other credential after 8+ months of college enrollment by Q24 (%)	+14.2***	3.1	+12.0***	2.9
Secondary Outcomes: Education (NSC)	**Full Sample**			
Received associate or higher degree by Q24 (%)	+10.2***	3.2	+8.5***	3.0
Total months with any college enrollment across Years 1-6	+4.7***	1.0	+4.1***	1.0
Total months with any full-time college enrollment across Years 1-6	+1.9***	0.6	+1.6***	0.6
Cumulative FTE months of college enrollment across Years 1-6	+3.3***	0.8	+2.8***	0.8
Secondary Outcomes: Financial Well-Being (Survey)	**Survey Respondents without Weights**			
Ability to handle financial emergency of $400 from savings or checking (%)	−1.1	3.6	−2.2	3.5
Average total unsecured debt ($)	−940	1,363	−1,340	1,407
Received means-tested public benefits (%)	−10.1***	3.7	−10.3***	3.7
Extent of financial distress (9-item scale, ranging from 0 to 9)	−0.3**	0.1	−0.2**	0.1
Selected Outcome: Financial Well-Being (Survey)	**Survey Respondents without Weights**			
Household income in month prior to interview ($)	+50	190	−8	185
Sample sizes (treatment + control group): NDNH 955 NSC 958 Survey 730				

Source: National Directory of New Hires (data received as of June 15, 2021). National Student Clearinghouse. PACE six-year follow-up survey.

Statistical significance levels for exploratory outcomes are based on two-tailed tests. For confirmatory and secondary outcomes, statistical significance levels are based on one-tailed *t*-tests tests of positive differences between research groups for positive outcomes and negative differences for negative outcomes (such as student debt). Statistical significance levels are summarized as follows: *** 1 percent level; ** 5 percent level; * 10 percent level.

Exhibit E-1d: Comparison of Confirmatory and Secondary Impact Estimates, Unadjusted and Adjusted for Baseline Imbalances – Year Up

Domain (Data Source), Outcome	Unadjusted Estimate		Adjusted Estimate	
	Impact	Standard Error	Impact	Standard Error
Confirmatory Outcome: Earnings/Employment (NDNH)	**Full Sample**			
Average quarterly earnings Q23-Q24 after randomization ($)	+1,933***	280	+1,881***	267
Selected Outcomes: Earnings/Employment (NDNH)	**Full Sample**			
Employed in Q23 after randomization (%)	+0.1	1.6	+0.0	1.6
Total earnings in Years 1-6 after randomization ($)	+30,212***	3,512	+29,865***	3,104
Secondary Outcomes: Employment (Survey)	**Survey Respondents without Weights**			
Working full-time (35+ hours/week) (%)	+7.8***	2.5	+7.8***	2.5
Working at job in a program target occupation (%)	+22.6***	2.4	+22.5***	2.3
Working at job providing all of a list of five possible benefits (%)	+11.1***	2.7	+10.6***	2.7
Access to career network (count of 6 yes/no items)	+0.2**	0.1	+0.2**	0.1
Selected Outcome: Education (NSC)	**Full Sample**			
Received a degree or some other credential after 8+ months of college enrollment by Q24 (%)	−1.5	1.9	−1.8	1.8
Secondary Outcomes: Financial Well-Being (Survey)	**Survey Respondents without Weights**			
Ability to handle financial emergency of $400 from savings or checking (%)	+8.2***	2.6	+7.2***	2.5
Average total unsecured debt ($)	−2,479***	870	−2,355***	872
Received means-tested public benefits (%)	−8.2***	2.6	−7.5***	2.3
Extent of financial distress (9-item scale, ranging from 0 to 9)	−0.1	0.1	−0.1	0.1
Selected Outcome: Financial Well-Being (Survey)	**Survey Respondents without Weights**			
Household income in month prior to interview ($)	+395***	151	+356**	140
Sample sizes (treatment + control group): NDNH 2,495 Survey 1,644				

Source: National Directory of New Hires (data received as of June 15, 2021). National Student Clearinghouse. PACE six-year follow-up survey.

Statistical significance levels for exploratory outcomes are based on two-tailed tests. For confirmatory and secondary outcomes, statistical significance levels are based on one-tailed *t*-tests tests of positive differences between research groups for positive outcomes and negative differences for negative outcomes (such as student debt). Statistical significance levels are summarized as follows: *** 1 percent level; ** 5 percent level; * 10 percent level.

E.1.2 Other PACE Sites

Exhibit E-1e: Comparison of Confirmatory and Secondary Impact Estimates, Unadjusted and Adjusted for Baseline Imbalances – Bridge to Employment in the Healthcare Industry

Domain (Data Source), Outcome	Unadjusted Estimate		Adjusted Estimate	
	Impact	Standard Error	Impact	Standard Error
Confirmatory Outcome: Earnings/Employment (NDNH)	**Full Sample**			
Average quarterly earnings Q23-Q24 after randomization ($)	−21	385	−134	358
Selected Outcomes: Earnings/Employment (NDNH)	**Full Sample**			
Employed in Q23 after randomization (%)	−1.0	2.8	−1.1	2.8
Total earnings in Years 1-6 after randomization ($)	+2,653	5,107	+1,208	4,455
Confirmatory Outcome: Education (NSC)	**Full Sample**			
Received a degree or some other credential after 8+ months of college enrollment by Q24 (%)	−2.5	1.9	−2.6	1.8
Sample sizes (treatment + control group): NDNH 974 NSC 1,004				

Source: National Directory of New Hires (data received as of June 15, 2021). National Student Clearinghouse. .

Statistical significance levels for exploratory outcomes are based on two-tailed tests. For confirmatory and secondary outcomes, statistical significance levels are based on one-tailed *t*-tests tests of positive differences between research groups for positive outcomes and negative differences for negative outcomes (such as student debt). Statistical significance levels are summarized as follows: *** 1 percent level; ** 5 percent level; * 10 percent level.

Exhibit E-1f: Comparison of Confirmatory and Secondary Impact Estimates, Unadjusted and Adjusted for Baseline Imbalances – Health Careers for All

Domain (Data Source), Outcome	Unadjusted Estimate		Adjusted Estimate	
	Impact	Standard Error	Impact	Standard Error
Confirmatory Outcome: Earnings/Employment (NDNH)	**Full Sample**			
Average quarterly earnings Q23-Q24 after randomization ($)	+640	655	+122	603
Selected Outcomes: Earnings/Employment (NDNH)	**Full Sample**			
Employed in Q23 after randomization (%)	−1.0	3.6	−2.6	3.6
Total earnings in Years 1-6 after randomization ($)	+7,354	7,994	−2,392	7,191
Selected Outcome: Education (NSC)	**Full Sample**			
Received a degree or some other credential after 8+ months of college enrollment by Q24 (%)	0.7	3.1	−1.7	2.9
Sample sizes (treatment + control group): NDNH 648 NSC 652				

Source: National Directory of New Hires (data received as of June 15, 2021). National Student Clearinghouse.

Statistical significance levels for exploratory outcomes are based on two-tailed tests. For confirmatory and secondary outcomes, statistical significance levels are based on one-tailed *t*-tests tests of positive differences between research groups for positive outcomes and negative differences for negative outcomes (such as student debt). Statistical significance levels are summarized as follows: *** 1 percent level; ** 5 percent level; * 10 percent level.

Exhibit E-1g: Comparison of Confirmatory and Secondary Impact Estimates, Unadjusted and Adjusted for Baseline Imbalances – Pathways to Healthcare

Domain (Data Source), Outcome	Unadjusted Estimate		Adjusted Estimate	
	Impact	Standard Error	Impact	Standard Error
Confirmatory Outcome: Earnings/Employment (NDNH)	**Full Sample**			
Average quarterly earnings Q23-Q24 after randomization ($)	−341	299	−330	283
Selected Outcomes: Earnings/Employment (NDNH)	**Full Sample**			
Employed in Q23 after randomization (%)	−1.1	2.8	−1.1	2.7
Total earnings in Years 1-6 after randomization ($)	−3,312	3,891	−3,244	3,338
Secondary Outcomes: Education (NSC)	**Full Sample**			
Received associate or higher degree by Q24 (%)	1.3	1.5	1.6	1.4
Received any college credential after Year 3 (%)	0.5	1.7	0.7	1.6
Enrolled in college sometime after Year 3 (%)	0.0	2.5	0.2	2.4
Sample sizes (treatment + control group): NDNH 1,208 NSC 1,217				

Source: National Directory of New Hires (data received as of June 15, 2021). National Student Clearinghouse.

Statistical significance levels for exploratory outcomes are based on two-tailed tests. For confirmatory and secondary outcomes, statistical significance levels are based on one-tailed *t*-tests tests of positive differences between research groups for positive outcomes and negative differences for negative outcomes (such as student debt). Statistical significance levels are summarized as follows: *** 1 percent level; ** 5 percent level; * 10 percent level.

Exhibit E-1h: Comparison of Confirmatory and Secondary Impact Estimates, Unadjusted and Adjusted for Baseline Imbalances – Patient Care Pathways Program

Domain (Data Source), Outcome	Unadjusted Estimate		Adjusted Estimate	
	Impact	Standard Error	Impact	Standard Error
Confirmatory Outcome: Earnings/Employment (NDNH)	**Full Sample**			
Average quarterly earnings Q23-Q24 after randomization ($)	−76	476	−182	450
Selected Outcomes: Earnings/Employment (NDNH)	**Full Sample**			
Employed in Q23 after randomization (%)	−0.3	3.4	−0.9	3.4
Total earnings in Years 1-6 after randomization ($)	−2,377	6,380	−6,028	5,218
Confirmatory Outcome: Education (NSC)	**Full Sample**			
Received a degree or some other credential after 8+ months of college enrollment by Q24 (%)	2.6**	1.0	2.2**	1.0
Secondary Outcome: Education (NSC)	**Full Sample**			
Received associate or higher degree by Q24 (%)	6.3**	3.4	6.8**	3.3
Sample sizes (treatment + control group): NDNH 486 NSC 499				

Source: National Directory of New Hires (data received as of June 15, 2021). National Student Clearinghouse.

Statistical significance levels for exploratory outcomes are based on two-tailed tests. For confirmatory and secondary outcomes, statistical significance levels are based on one-tailed *t*-tests tests of positive differences between research groups for positive outcomes and negative differences for negative outcomes (such as student debt). Statistical significance levels are summarized as follows: *** 1 percent level; ** 5 percent level; * 10 percent level.

Exhibit E-1i: Comparison of Confirmatory and Secondary Impact Estimates, Unadjusted and Adjusted for Baseline Imbalances – Workforce Training Academy Connect

Domain (Data Source), Outcome	Unadjusted Estimate		Adjusted Estimate	
	Impact	Standard Error	Impact	Standard Error
Confirmatory Outcome: Earnings/Employment (NDNH)	**Full Sample**			
Average quarterly earnings Q23-Q24 after randomization ($)	−21	316	−157	282
Selected Outcomes: Earnings/Employment (NDNH)	**Full Sample**			
Employed in Q23 after randomization (%)	+4.1	3.2	+2.8	3.1
Total earnings in Years 1-6 after randomization ($)	+1,605	4,888	−1,834	3,754
Selected Outcome: Education (NSC)	**Full Sample**			
Received a degree or some other credential after 8+ months of college enrollment by Q24 (%)	2.6**	1.0	2.2**	1.0
Sample sizes (treatment + control group): NDNH 920 NSC 943				

Source: National Directory of New Hires (data received as of June 15, 2021). National Student Clearinghouse.

Statistical significance levels for exploratory outcomes are based on two-tailed tests. For confirmatory and secondary outcomes, statistical significance levels are based on one-tailed *t*-tests tests of positive differences between research groups for positive outcomes and negative differences for negative outcomes (such as student debt). Statistical significance levels are summarized as follows: *** 1 percent level; ** 5 percent level; * 10 percent level.

E.2 Nonresponse Weighting Sensitivity Analysis

Exhibits E-2a to E-2d present evidence about the level of nonresponse bias with and without adjustment weights. The first set of impact estimates (column 1), which is available only for NDNH- and NSC-based estimates, is based on the full sample. The second set of impact estimates (column 3) excludes survey nonrespondents. Differences between the first and second set of impacts signal nonresponse bias. The third set of impact estimates (column 5) also excludes survey nonrespondents but weights survey respondents with nonresponse-adjustment weights, which are explained in Appendix Section B.4. If the weights are good, then the differences between the first and fifth columns will be smaller than those between the first and third columns. Note that all three sets of impact estimates are regression-adjusted with the covariates discussed in Appendix Section A.2. We did not formally test the differences between the alternative estimates, but given that the survey respondents constitute a very large subset of all participants for each of the four surveyed sites, many of the differences would be statistically significant.

One potential disadvantage of weighting to adjust for survey nonresponse is the potential for variance inflation, causing a loss of power. This does not appear to be a serious problem in any of the four surveyed sites. When comparing estimates on the survey respondent sample, the weighted standard errors are only 1.7, 0.1, and 1.3 percent larger on average than the unweighted standard errors for I-BEST, VIDA, and Year Up, respectively. Larger standard errors are not universal. Somewhat surprisingly, standard errors on the weighted impacts at Carreras are 0.8 percent smaller on average than their unweighted counterparts.

When considering impact bias, the nonresponse weights performed well for Carreras. Weighting reduced bias for three of the four analyzed NDNH- and NSC-based impacts, with the lone exception being average quarterly earnings Q23-Q24 after randomization; however, in this case both the weighted and unweighted biases are small. The effect of weighting on survey-based outcomes appears small and does not affect the significance of any key outcomes.

Exhibit E-2a: Comparison of Confirmatory and Secondary Estimates of the Impact of Carreras en Salud for the Unweighted and Weighted Survey Samples

	Full Sample		Unweighted Sample		Weighted Sample	
Outcome (Data Source)	Impact Estimate	Standard Error	Impact Estimate	Standard Error	Impact Estimate	Standard Error
Confirmatory Outcome: Earnings/Employment (NDNH)						
Average quarterly earnings Q23-Q24 after randomization ($)	+360	374	+394	481	+304	461
Selected Outcomes: Earnings/Employment (NDNH)						
Employed in Q23 after randomization (%)	+ 4.4*	3.1	+1.2	3.8	+1.5	3.9
Total earnings in Years 1-6 after randomization ($)	−4,259	4,680	−5,641	5,790	−4,273	5,770
Secondary Employment Outcomes (Survey)						
Working full-time (35+ hours/week) (%)			+1.2	4.3	+1.2	4.4
Working at job in a program target occupation (%)			+5.9*	3.9	+5.7*	3.9
Working at job providing all of a list of five possible benefits (%)			+0.2	4.3	−0.1	4.4
Access to career network (count of 6 yes/no items)			+0.1	0.2	+0.1	0.2
Confirmatory Outcome: Education (NSC)						
Received a degree or some other credential after 8+ months of college enrollment by Q24 (%)	+3.8*	2.4	+5.1*	3.1	+3.8*	2.9
Secondary Outcomes: Financial Well-Being (Survey)						
Ability to handle financial emergency of $400 from savings or checking (%)			−9.7	4.1	−9.5	4.1
Average total unsecured debt ($)			−147	1,646	−90	1,551
Received means-tested public benefits (%)			+0.7	4.3	−1.0	4.3
Extent of financial distress (9-item scale, ranging from 0 to 9)			−0.1	0.1	−0.1	0.1
Sample sizes (treatment + control groups):						
NDNH	775		520		520	
NSC	799		536		536	
Survey	na		536		536	

Source: National Directory of New Hires (data received as of March 16, 2021). National Student Clearinghouse. PACE six-year follow-up survey.

Note: All estimates are regression-adjusted as discussed in Appendix Section A.2. The full sample columns are blank for survey-measured outcomes because they are not available for the full sample. The weighted NDNH-based figures presented in this exhibit use an approximation of the final nonresponse weights. This approximation was prepared strictly in terms of data available on ACF laptops.

Statistical significance levels for exploratory outcomes are based on two-tailed tests. For confirmatory and secondary outcomes, statistical significance levels are based on one-tailed *t*-tests tests of positive differences between research groups for positive outcomes and negative differences for negative outcomes (such as student debt). Statistical significance levels are summarized as follows: *** 1 percent level; ** 5 percent level; * 10 percent level.

Though weighting reduced bias for Carreras estimates, this does not appear to be the case for I-BEST estimates, as weighting increased bias for each of the NDNH-based outcomes (although slightly reducing bias for the NSC-based confirmatory outcome).[81] That nonresponse

81 As discussed in Section B.4.3.

weights performed worse for I-BEST than for the rest of the PACE studies may not be surprising given the comparatively low response rate and sample size at I-BEST.[82] Weighting did not appear to aid the impact estimates for the I-BEST study, but the benefits observed in other studies support the decision to implement nonresponse weighting for the PACE sample.

Exhibit E-2b: Comparison of Confirmatory and Secondary Estimates of the Impact of I-BEST for the Unweighted and Weighted Survey Samples

	Full Sample		Unweighted Sample		Weighted Sample	
Outcome (Data Source)	Impact Estimate	Standard Error	Impact Estimate	Standard Error	Impact Estimate	Standard Error
Confirmatory Outcome: Earnings/Employment (NDNH)						
Average quarterly earnings Q23-Q24 after randomization ($)	+123	450	−164	603	−440	602
Selected Outcomes: Earnings/Employment (NDNH)						
Employed in Q23 after randomization (%)	+3.2	3.9	+3.2	4.9	+0.1	5.3
Total earnings in Years 1-6 after randomization ($)	4,061	5,891	2,598	7,926	925	8,244
Secondary Employment Outcomes (Survey)						
Working full-time (35+ hours/week) (%)			+6.2	5.2	+3.0	5.3
Working at job in a program target occupation (%)			+8.8***	3.2	+7.5***	3.2
Working at job providing all of a list of five possible benefits (%)			+3.6	5.1	+2.4	5.1
Access to career network (count of 6 yes/no items)			+0.1	0.2	+0.0	0.2
Confirmatory Outcome: Education (NSC)						
Received a degree or some other credential after 8+ months of college enrollment by Q24 (%)	+1.1	2.6	+2.8	3.8	+0.2	4.0
Secondary Outcomes: Financial Well-Being (Survey)						
Ability to handle financial emergency of $400 from savings or checking (%)			+7.1*	5.2	+5.9	5.5
Average total unsecured debt ($)			+3,425	6,036	+3,953	5,612
Received means-tested public benefits (%)			+2.6	5.0	+1.5	5.2
Extent of financial distress (9-item scale, ranging from 0 to 9)			−0.2*	0.2	−0.1	0.2
Sample sizes (treatment + control groups):						
NDNH	610		352		352	
NSC	631		358		358	
Survey	na		358		358	

Source: NDNH (data received as of March 16, 2021). National Student Clearinghouse. PACE six-year follow-up survey.

Note: All estimates are regression-adjusted as discussed in Appendix Section A.2. The full sample columns are blank for survey-measured outcomes because they are not available for the full sample. The weighted NDNH-based figures presented in this exhibit use an approximation of the final nonresponse weights. This approximation was prepared strictly in terms of data available on ACF laptops. Statistical significance levels for exploratory outcomes are based on two-tailed tests. For confirmatory and secondary outcomes, statistical significance levels are based on one-tailed *t*-tests tests of positive differences between research groups for positive outcomes and negative differences for negative outcomes (such as student debt). Statistical significance levels are summarized as follows: *** 1 percent level; ** 5 percent level; * 10 percent level.

82 I-BEST had an overall response rate of 58.3 percent, notably lower than the next lowest site, Year Up, at 65.7 percent and far lower than VIDA, which had the highest overall response rate at 77.3 percent.

Exhibit E-2c: Comparison of Confirmatory and Secondary Estimates of the Impact of VIDA for the Unweighted and Weighted Survey Samples

Outcome (Data Source)	Full Sample		Unweighted Sample		Weighted Sample	
	Impact Estimate	Standard Error	Impact Estimate	Standard Error	Impact Estimate	Standard Error
Confirmatory Outcome: Earnings/Employment (NDNH)						
Average quarterly earnings Q23-Q24 after randomization ($)	+28	426	+424	493	+105	496
Selected Outcomes: Earnings/Employment (NDNH)						
Employed in Q23 after randomization (%)	+1.0	2.6	+3.6	3.0	+2.5	3.0
Total earnings in Years 1-6 after randomization ($)	−5,784	5,764	+396	6,653	−3,934	6,707
Secondary Employment Outcomes (Survey)						
Working full-time (35+ hours/week) (%)			+0.6	3.5	−1.3	3.5
Working at job providing all of a list of five possible benefits (%)			+4.1	3.8	+2.2	3.8
Access to career network (count of 6 yes/no items)			+0.3**	0.2	+0.3**	0.2
Confirmatory Outcome: Education (NSC)						
Received a degree or some other credential after 8+ months of college enrollment by Q24 (%)	+12.0***	2.9	+11.3***	3.3	+11.2***	3.4
Secondary Outcomes: Education (NSC)						
Received AA or higher degree by Q24 (%)	+8.5***	3.0	+8.4***	3.6	+8.1**	3.5
Total months with any college enrollment across Years 1-6	+4.1***	1.0	+3.3***	1.2	+4.0***	1.2
Total months with any full-time college enrollment across Years 1-6	+1.6***	0.6	+1.4**	0.7	+1.7***	0.7
Cumulative FTE months of college enrollment across Years 1-6	+2.8***	0.8	+2.3***	0.9	+2.8***	0.9
Secondary Outcomes: Financial Well-Being (Survey)						
Ability to handle financial emergency of $400 from savings or checking (%)			−2.2	3.5	−2.8	3.5
Average total unsecured debt ($)			−1,340	1,407	−1,158	1,395
Received means-tested public benefits (%)			−10.3***	3.7	−8.5**	3.7
Extent of financial distress (9-item scale, ranging from 0 to 9)			−0.2**	0.1	−0.2*	0.1
Sample sizes (treatment + control groups):						
NDNH	955		729		729	
NSC	958		732		732	
Survey	na		732		732	

Source: National Directory of New Hires (data received as of March 16, 2021). National Student Clearinghouse. PACE six-year follow-up survey.

Note: All estimates are regression-adjusted as discussed in Appendix Section A.2. The full sample columns are blank for survey-measured outcomes because they are not available for the full sample.

The weighted NDNH-based figures presented in this exhibit use an approximation of the final nonresponse weights. This approximation was prepared strictly in terms of data available on ACF laptops.

Statistical significance levels for exploratory outcomes are based on two-tailed tests. For confirmatory and secondary outcomes, statistical significance levels are based on one-tailed *t*-tests tests of positive differences between research groups for positive outcomes and negative differences for negative outcomes (such as student debt). Statistical significance levels are summarized as follows: *** 1 percent level; ** 5 percent level; * 10 percent level.

The decision to adjust for survey nonresponse did not affect the significance of any key outcomes in the VIDA study, but we still observe noticeable differences in the unweighted survey sample and full sample with respect to the NDNH-based impacts. Nonresponse weighting performed well in mitigating these differences, particularly in the earnings impacts. Additionally, weighting reduced nonresponse bias for three of the five key NSC-based impact estimates.

We observe bias reduction in all three key NDNH outcomes for Year Up. In the case of Q23 employment, nonresponse weighting appears to remove virtually all nonresponse bias. Despite the differences in point estimates among the three columns, all estimates tell the same overall story for Year Up—that the intervention had significant and substantial career benefits for its participants who found employment.

Exhibit E-2d: Comparison of Confirmatory and Secondary Estimates of the Impact of Year Up for the Unweighted and Weighted Survey Samples

	Full Sample		Unweighted Sample		Weighted Sample	
Outcome (Data Source)	Impact Estimate	Standard Error	Impact Estimate	Standard Error	Impact Estimate	Standard Error
Confirmatory Outcome: Earnings/Employment (NDNH)						
Average quarterly earnings Q23-Q24 after randomization ($)	+1,893***	269	+2,303***	341	+2,039***	337
Selected Outcomes: Earnings/Employment (NDNH)						
Employed in Q23 after randomization (%)	+0.0	1.6	+0.8	2.0	+0.0	2.0
Total earnings in Years 1-6 after randomization ($)	29,665***	3,127	35,029***	3,895	31,414***	3,894
Secondary Employment Outcomes (Survey)						
Working full-time (35+ hours/week) (%)			+7.8***	2.5	+5.2**	2.5
Working at job in a program target occupation (%)			+22.5***	2.3	+21.3***	2.4
Working at job providing all of a list of five possible benefits (%)			+10.6***	2.7	+9.2***	2.8
Access to career network (count of 6 yes/no items)			+0.2**	0.1	+0.2**	0.1
Secondary Outcomes: Financial Well-Being (Survey)						
Ability to handle financial emergency of $400 from savings or checking (%)			+7.2***	2.5	+6.3***	2.5
Average total unsecured debt ($)			−2,355***	872	−2,231***	899
Received means-tested public benefits (%)			−7.5***	2.3	−7.0***	2.4
Extent of financial distress (9-item scale, ranging from 0 to 9)			−0.1	0.1	−0.1	0.1
Sample sizes (treatment + control groups):						
NDNH	2,495		1,622		1,622	
NSC	2,539		1,653		1,653	
Survey	na		1,653		1,653	

Source: National Directory of New Hires (data received as of March 16, 2021). PACE six-year follow-up survey.

Note: All estimates are regression-adjusted as discussed in Appendix Section A.2. The full sample columns are blank for survey-measured outcomes because they are not available for the full sample. The weighted NDNH-based figures presented in this exhibit use an approximation of the final nonresponse weights. This approximation was prepared strictly in terms of data available on ACF laptops.

Statistical significance levels for exploratory outcomes are based on two-tailed tests. For confirmatory and secondary outcomes, statistical significance levels are based on one-tailed *t*-tests tests of positive differences between research groups for positive outcomes and negative differences for negative outcomes (such as student debt). Statistical significance levels are summarized as follows: *** 1 percent level; ** 5 percent level; * 10 percent level.

E.3 Imputation Sensitivity Analysis

Specific details regarding imputation of income and debt outcomes are discussed in Appendix Section B.3. Exhibits E-3a to E3d below present site estimates for household income, personal income, and student debt with and without imputation in order to gauge the sensitivity of estimated impacts to the imputation procedures, as well as to document any associated improvements in precision. The exhibits below show that most of the inferences on impacts are robust to the use of imputation. There are three minor exceptions. For each of Carreras and VIDA, imputation removed a star, while at Year Up, imputation added a single star. The differences between impacts estimated with and without imputation are never statistically significant. As expected, most of the standard errors with imputation are smaller than those without imputation, but the differences are mild. This is a sign that the multiple imputation is working as intended to mitigate the negative bias in standard errors so often seen in estimates based on imputed data.

Exhibit E-3a: Impacts of Carreras en Salud on Select Survey Outcomes with and without Imputation

Outcome	Treatment	Control	Impact	Standard Error	Total Sample Size
Household income					
With imputation ($)	42,641	45,479	−2,838	2,322	536
Without imputation ($)	42,790	46,935	−4,145*	2,443	436
Difference	−150	−1,456	1,307	1,258	
Personal income					
With imputation ($)	21,528	22,799	−1,271	1,416	536
Without imputation ($)	22,061	22,890	−829	1,429	507
Difference	−533	−91	−442	391	
Student debt in student's name					
With imputation ($)	2,154	2,898	−744	610	536
Without imputation ($)	2,166	2,887	−721	614	532
Difference	−12	11	−23	36	

Source: PACE six-year follow-up survey.
Note: Statistically significant in a two-tailed test as follows: *** 1 percent level; ** 5 percent level; * 10 percent level. Statistical significance levels for two-sided tests of differences in impacts are indicated with dagger, as follows: †††=1 percent, ††=5 percent, †=10 percent.

Exhibit E-3b: Impacts of I-BEST on Select Survey Outcomes with and without Imputation

Outcome	Treatment	Control	Impact	Standard Error	Total Sample Size
Household income					
With imputation ($)	48,038	45,003	+3,035	3,650	358
Without imputation ($)	48,377	45,126	+3,250	3,793	323
Difference	−339	−123	−215	1,284	
Personal income					
With imputation ($)	27,011	25,642	+1,369	2,650	358
Without imputation ($)	27,045	25,748	+1,296	2,666	350
Difference	−33	−106	+73	449	
Student debt in student's name					
With imputation ($)	2,805	2,661	+145	892	358
Without imputation ($)	2,835	2,682	+154	896	355
Difference	−30	−21	−9	24	

Source: PACE six-year follow-up survey.

Note: Statistically significant in a two-tailed test as follows: *** 1 percent level; ** 5 percent level; * 10 percent level. Statistical significance levels for two-sided tests of differences in impacts are indicated with dagger, as follows: †††=1 percent, ††=5 percent, †=10 percent.

Exhibit E-3c: Impacts of VIDA on Select Survey Outcomes with and without Imputation

Outcome	Treatment	Control	Impact	Standard Error	Total Sample Size
Household income					
With imputation ($)	52,295	54,007	−1,711	2,408	732
Without imputation ($)	51,367	52,989	−1,622	2,397	637
Difference	+929	+1,018	−89	1,226	
Personal income					
With imputation ($)	33,178	33,099	+79	1,749	732
Without imputation ($)	33,369	32,972	+397	1,774	695
Difference	−191	+127	−318	463	
Student debt in student's name					
With imputation ($)	4,574	5,844	−1,270	772	732
Without imputation ($)	4,569	5,889	−1,320*	779	723
Difference	+5	−45	+50	42	

Source: PACE six-year follow-up survey.

Note: Statistically significant in a two-tailed test as follows: *** 1 percent level; ** 5 percent level; * 10 percent level. Statistical significance levels for two-sided tests of differences in impacts are indicated with dagger, as follows: †††=1 percent, ††=5 percent, †=10 percent.

Exhibit E-3d: Impacts of Year Up on Select Survey Outcomes with and without Imputation

Outcome	Treatment	Control	Impact	Standard Error	Total Sample Size
Household income					
With imputation ($)	58,760	55,596	+3,164*	1,781	1,653
Without imputation ($)	56,828	55,460	+1,368	1,943	1,248
Difference	+1,932	+136	+1,796	1,176	
Personal income					
With imputation ($)	35,708	30,333	+5,374***	1,143	1,653
Without imputation ($)	35,699	30,283	+5,416***	1,148	1,546
Difference	+9	+50	−42	399	
Student debt in student's name					
With imputation ($)	2,983	4,920	−1,936***	500	1,653
Without imputation ($)	2,861	4,707	−1,846***	506	1,600
Difference	+122	+213	−91	123	

Source: PACE six-year follow-up survey.

Note: Statistically significant in a two-tailed test as follows: *** 1 percent level; ** 5 percent level; * 10 percent level. Statistical significance levels for two-sided tests of differences in impacts are indicated with dagger, as follows: †††=1 percent, ††=5 percent, †=10 percent.

E.4 NSC Sensitivity Analysis

The exhibits below display NSC- and survey-based estimates for degree receipt, college credential receipt, and current enrollment obtained. NSC records generally underestimate receipt of any credential (including both degrees and sub-degree credentials) compared to survey records. Part of this is probably due to some colleges not participating in NSC; in general, however, the causes of the differences are not known. Additionally, NSC records under-report degree receipt compared to self-reports, albeit to a lesser degree than seen for any college-issued credential.

Despite these differences in treatment and control group estimates, inferences about program impacts are generally robust to the choice of data source. The one clear exception is current college enrollment at VIDA. The survey shows a highly significant impact on this outcome, whereas the NSC shows only a small and statistically insignificant impact. The difference between the two estimated impacts is itself statistically significant. Research into this discrepancy failed to suggest any clear reasons.[83] Other and more minor exceptions include the gain or loss of stars without the difference between the alternative impacts reaching the threshold for statistical significance. This occurred at I-BEST (1 versus 0 stars for earned degrees), VIDA (3 versus 2 stars for earned degrees), and Year Up (0 versus 2 stars for earned degrees and 3 versus 0 stars on any credentials).

Regarding the inconsistencies for Year Up, a similar pattern appeared in its Intermediate Outcomes Study report (Fein and Dastrup 2021). One potential explanation for these discrepancies is confusion with respect to college enrollment. As discussed in the appendices to that report (Judkins, Walton, Durham et al. 2021), NSC estimates of college enrollment in the

[83] To study this discrepancy, we looked at the study participants for whom the survey-based "Enrolled as of the survey" measure differed from the NSC-based "Q24 enrollment measure"; however, there was not a sufficient number of such cases for any clear patterns to emerge.

early quarters following randomization (i.e., the quarters in which Year Up training occurred) were significantly higher for participants in the treatment group than were survey estimates for those participants. However, this discrepancy did not exist for later quarters (i.e., quarters after Year Up training would have concluded) at three years. Furthermore, estimates of college enrollment among the Year Up control group did not show major differences at three years. This suggested that many of those receiving Year Up training did not view their training as college enrollment, even though they were registered at Year Up partner local community colleges. Therefore, it is possible that many of those who received credentials through Year Up training did not view their credentials as earned "from a college,"[84] whereas the local Year Up partner college reported their enrollment and credentials to the NSC. Because this discrepancy would be localized to the Year Up control group, it could dampen impacts for credential receipt.

Exhibit E-4a: Impacts of Carreras en Salud on Educational Progress Six Years after Random Assignment Based on NSC Records and Self-Reports

Outcome	Treatment	Control	Impact	Standard Error
Any Degrees				
Any NSC-reported degrees (%)	10.7	8.8	+1.9	2.0
Any self-reported degrees (%)	13.9	14.8	−0.9	2.9
Difference	−3.2	−6.0	+2.9	2.7
Any Completions from a College				
Any NSC completions through Q24 (%)	21.3	16.1	+5.2**	2.6
Any credentials from a college as of survey (%)	37.7	29.3	+8.4**	4.0
Difference	−16.4	−13.2	−3.2	4.0
Current College Enrollment				
Enrolled during any part of Q24 (%)	10.1	8.8	+1.3	2.1
Enrolled as of the survey (%)	9.7	11.9	−2.3	2.8
Difference	+0.4	−3.1	+3.5	2.3
Sample sizes (treatment + control group): NSC 799 Survey 536				

Source: National Student Clearinghouse. PACE six-year follow-up survey.

Note: A majority of survey interviews occurred in Q23 and Q24 after randomization.

Statistically significant in a two-tailed test as follows: *** 1 percent level; ** 5 percent level; * 10 percent level. Statistical significance levels for two-sided tests of differences in impacts are indicated with dagger, as follows: †††=1 percent, ††=5 percent, †=10 percent.

84 The survey instrument asked respondents whether they had earned any degrees for completing "any regular college classes" and/or any credentials "for completing a program at a community or technical college." If Year Up training group members did not understand that their Year Up training counted as college enrollment, they may also have been less likely to believe any credentials earned were through a college.

Exhibit E-4b: Impacts of I-BEST on Educational Progress Six Years after Random Assignment Based on NSC Records and Self-Reports

Outcome	Treatment	Control	Impact	Standard Error
Any Degrees				
Any NSC-reported degrees (%)	10.7	7.0	+3.7*	2.2
Any self-reported degrees (%)	10.8	8.2	+2.6	3.3
Difference	−0.1	−1.2	+1.1	2.8
Any Completions from a College				
Any NSC completions through Q24 (%)	28.2	14.2	+14.0***	3.2
Any credentials from a college as of survey (%)	44.0	28.2	+15.8***	5.3
Difference	−15.8	−14.0	−1.8	5.3
Current College Enrollment				
Enrolled during any part of Q24 (%)	7.0	5.1	+2.0	1.9
Enrolled as of the survey (%)	7.9	9.0	−1.1	3.3
Difference	−0.9	−3.9	+3.1	3.0
Sample sizes (treatment + control group): NSC 631 Survey 356				

Source: National Student Clearinghouse. PACE six-year follow-up survey.

Note: A majority of survey interviews occurred in Q23 and Q24 after randomization.

Statistically significant in a two-tailed test as follows: *** 1 percent level; ** 5 percent level; * 10 percent level. Statistical significance levels for two-sided tests of differences in impacts are indicated with dagger, as follows: †††=1 percent, ††=5 percent, †=10 percent.

Exhibit E-4c: Impacts of VIDA on Educational Progress Six Years after Random Assignment Based on NSC Records and Self-Reports

Outcome	Treatment	Control	Impact	Standard Error
Any Degrees				
Any NSC-reported degrees (%)	48.9	40.4	+8.5***	3.0
Any self-reported degrees (%)	49.2	41.7	+7.5**	3.5
Difference	−0.3	−1.3	+1.0	3.0
Any Completions from a College				
Any NSC completions through Q24 (%)	72.3	58.8	+13.6***	2.8
Any credentials from a college as of survey (%)	69.6	60.2	+9.4***	3.4
Difference	2.7	−1.4	+4.2	3.3
Current College Enrollment				
Enrolled during any part of Q24 (%)	14.6	12.9	+1.6	2.2
Enrolled as of the survey (%)	17.0	9.8	+7.2***	2.5
Difference	−2.4	+3.1	−5.6†††	2.1
Sample sizes (treatment + control group): NSC 958 Survey 730				

Source: National Student Clearinghouse. PACE six-year follow-up survey.

Note: A majority of survey interviews occurred in Q23 and Q24 after randomization.

Statistically significant in a two-tailed test as follows: *** 1 percent level; ** 5 percent level; * 10 percent level. Statistical significance levels for two-sided tests of differences in impacts are indicated with dagger, as follows: †††=1 percent, ††=5 percent, †=10 percent.

Exhibit E-4d: Impacts of Year Up on Educational Progress Six Years after Random Assignment Based on NSC Records and Self-Reports

Outcome	Treatment	Control	Impact	Standard Error
Any Degrees				
Any NSC-reported degrees (%)	9.5	11.3	−1.8	1.2
Any self-reported degrees (%)	11.6	15.4	−3.8**	1.8
Difference	−2.1	−4.1	+2.0	1.5
Any Completions from a College				
Any NSC completions through Q24 (%)	17.5	13.5	+4.0***	1.4
Any credentials from a college as of survey (%)	25.4	23.6	+1.8	2.3
Difference	−7.9	−10.1	+2.2	2.2
Current College Enrollment				
Enrolled during any part of Q24 (%)	12.5	12.2	+0.4	1.4
Enrolled as of the survey (%)	13.5	12.3	+1.2	1.7
Difference	−1.0	−0.1	−0.9	1.6
Sample sizes (treatment + control group): NSC 2,539 Survey 1,644				

Source: National Student Clearinghouse. PACE six-year follow-up survey.

Note: A majority of survey interviews occurred in Q23 and Q24 after randomization.

Statistically significant in a two-tailed test as follows: *** 1 percent level; ** 5 percent level; * 10 percent level. Statistical significance levels for two-sided tests of differences in impacts are indicated with dagger, as follows: †††=1 percent, ††=5 percent, †=10 percent.

E.5 NDNH Sensitivity Analysis

As discussed in Appendix D, we used data from the NDNH to analyze earnings and employment outcomes for the participants in the nine PACE studies. As a result, for the four surveyed sites, we have earnings and employment impact estimates both from the survey and from NDNH.[85] The four exhibits below compare these two data sources with respect to their estimates for impacts on Q24 earnings and employment. The first three exhibits show that these estimates are remarkably consistent for participants of Carreras, I-BEST, and VIDA; however, the two estimated impacts for earnings in Q24 after randomization are significantly different for Year Up.[86] Despite these differences, both sources show that Year Up treatment had substantial benefit for its participants, as both impacts are highly significant.

85 The six-year instrument asked about wages and hours at (one) current job if employed. We multiplied the implied weekly wages by 13 to get an approximation of quarterly earnings. Clearly, this will not be very accurate for people with sporadic employment or multiple jobs.

86 Possible reasons for this discrepancy are discussed further in the Year Up report (Fein forthcoming).

Exhibit E-5a: Impacts of Carreras en Salud on Earnings and Employment Six Years after Random Assignment Based on NDNH Records and Self-Reports

Outcome	Treatment	Control	Impact	Standard Error
Quarterly Earnings				
Average NDNH earnings in Q24 ($)	6,255	5,961	+294	370
Self-reported earnings as of survey ($)	6,528	6,198	+330	395
Difference	−273	−237	−36	350
Employment				
Percentage with employer-reported wages in Q24 (%)	75.4	73.2	+2.3	3.1
Percentage working in the week prior to survey interview (%)	78.7	75.7	+3.1	3.7
Difference	−3.3	−2.5	−0.8	3.5
Sample sizes (treatment + control group): NDNH 775 Survey 520				

Source: National Directory of New Hires (data received as of June 15, 2021). PACE six-year follow-up survey.

Note: A majority of survey interviews occurred in Q23 and Q24 after randomization. All estimates in this exhibit are restricted to individuals with social security numbers that OCSE deemed to be of sufficient quality for its matching purposes, as discussed in Appendix D.

Statistically significant in a two-tailed test as follows: *** 1 percent level; ** 5 percent level; * 10 percent level. Statistical significance levels for two-sided tests of differences in impacts are indicated with dagger, as follows: †††=1 percent, ††=5 percent, †=10 percent.

Exhibit E-5b: Impacts of I-BEST on Earnings and Employment Six Years after Random Assignment Based on NDNH Records and Self-Reports

Outcome	Treatment	Control	Impact	Standard Error
Quarterly Earnings				
Average NDNH earnings in Q24 ($)	5,375	5,062	+313	463
Self-reported earnings as of survey ($)	6,646	6,184	+462	717
Difference	−1,271	−1,122	−149	665
Employment				
Percentage with employer-reported wages in Q24 (%)	66.7	60.7	+6.0	3.8
Percentage working in the week prior to survey interview (%)	70.0	63.8	+6.2	5.3
Difference	−3.3	−2.5	−0.2	5.4
Sample sizes (treatment + control group): NDNH 610 Survey 352				

Source: National Directory of New Hires (data received as of June 15, 2021). PACE six-year follow-up survey.

Note: A majority of survey interviews occurred in Q23 and Q24 after randomization. All estimates in this exhibit are restricted to individuals with social security numbers that OCSE deemed to be of sufficient quality for its matching purposes, as discussed in Appendix D.

Statistically significant in a two-tailed test as follows: *** 1 percent level; ** 5 percent level; * 10 percent level. Statistical significance levels for two-sided tests of differences in impacts are indicated with dagger, as follows: †††=1 percent, ††=5 percent, †=10 percent.

Exhibit E-5c: Impacts of VIDA on Earnings and Employment Six Years after Random Assignment Based on NDNH Records and Self-Reports

Outcome	Treatment	Control	Impact	Standard Error
Quarterly Earnings				
Average NDNH earnings in Q24 ($)	8,369	8,361	+8	433
Self-reported earnings as of survey ($)	8,604	8,375	+229	475
Difference	−235	−14	−221	429
Employment				
Percentage with employer-reported wages in Q24 (%)	80.2	80.0	+0.3	2.5
Percentage working in the week prior to survey interview (%)	82.4	80.7	+1.7	2.9
Difference	−2.2	−0.7	−1.5	2.8
Sample sizes (treatment + control group): NDNH 955 Survey 729				

Source: National Directory of New Hires (data received as of June 15, 2021). PACE six-year follow-up survey.

Note: A majority of survey interviews occurred in Q23 and Q24 after randomization. All estimates in this exhibit are restricted to individuals with social security numbers that OCSE deemed to be of sufficient quality for its matching purposes, as discussed in Appendix D.

Statistically significant in a two-tailed test as follows: *** 1 percent level; ** 5 percent level; * 10 percent level. Statistical significance levels for two-sided tests of differences in impacts are indicated with dagger, as follows: †††=1 percent, ††=5 percent, †=10 percent.

Exhibit E-5d: Impacts of Year Up on Earnings and Employment Six Years after Random Assignment Based on NDNH Records and Self-Reports

Outcome	Treatment	Control	Impact	Standard Error
Quarterly Earnings				
Average NDNH earnings in Q24 ($)	8,714	6,846	+1,868***	282
Self-reported earnings as of survey ($)	9,145	7,902	+1,243***	323
Difference	−431	−6,201	+625††	314
Employment				
Percentage with employer-reported wages in Q24 (%)	78.6	78.4	+0.2	1.7
Percentage working in the week prior to survey interview (%)	80.0	81.1	−1.1	2.1
Difference	−1.4	−2.7	+1.3	2.3
Sample sizes (treatment + control group): NDNH 2,495 Survey 1,622				

Source: National Directory of New Hires (data received as of June 15, 2021). PACE six-year follow-up survey.

Note: A majority of survey interviews occurred in Q23 and Q24 after randomization. All estimates in this exhibit are restricted to individuals with social security numbers that OCSE deemed to be of sufficient quality for its matching purposes, as discussed in Appendix D.

Statistically significant in a two-tailed test as follows: *** 1 percent level; ** 5 percent level; * 10 percent level. Statistical significance levels for two-sided tests of differences in impacts are indicated with dagger, as follows: †††=1 percent, ††=5 percent, †=10 percent.

Appendix F: Treatment of Outliers

We took a conservative approach to outliers, retaining extreme values except where they were clearly impossible. This approach is based on the general difficulty of discriminating between errors and legitimate large values and on the fact that remedies require assumptions about true values that may not be correct.

Trimming observations could easily introduce non-ignorable nonresponse by making nonresponse a function of Y.[87]

Winsorizing observations (also known as "top-coding," where values above a threshold are set equal to the threshold) could introduce bias if there is a treatment impact but the same threshold is used for treatment and control group members (and there is no reasonable basis for setting different thresholds for the two groups).

Furthermore, evidence suggests that results are generally robust to extreme values. In particular, research by Judkins and Porter (2016) and Lumley et al. (2002) indicates that for the sample sizes available in this evaluation, ordinary least squares inference on the reported data should be robust to outliers.

Outcomes assessed for extreme values included instructional hours (by type of instruction), credits, and National Directory of New Hires earnings. We found one value that was clearly impossible, and therefore discarded data from the case.

87 Trimming by definition creates item nonresponse because the provided response is discarded. If trimming is a function of observed Y, as is standard, and if there is some relationship between observed Y and true Y, then item nonresponse becomes a function of true Y, which is known as "non-ignorable nonresponse." Because there is no known way to remove bias due to non-ignorable nonresponse, trimming is likely to create uncorrectable biases in estimated treatment effects.

References

Abt Associates Inc. 2014. *Pathways for Advancing Careers and Education (PACE). Evaluation Design Report.* OPRE Report 2014-76. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/opre/report/pathways-advancing-careers-and-education-evaluation-design-report.

Abt Associates Inc. 2015. *Pathways for Advancing Careers and Education (PACE). Technical Supplement to the Evaluation Design Report: Impact Analysis Plan.* OPRE Report 2015-100. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/opre/resource/pathways-for-advancing-careers-and-education-supplement-evaluation-design-impact-analysis-plan.

Bell, S. H., and L. L. Orr. 1994. "Is Subsidized Employment Cost Effective for Welfare Recipients? Experimental Evidence from Seven State Demonstrations." *Journal of Human Resources* 29 (1): 42-61.

Bühlmann, P., and S. van de Geer. 2011. *Statistics for High-Dimensional Data.* Berlin, Heidelberg, Germany: Springer.

Chen, Q., J. Zobel, X. Zhang, and K. Verspoor. 2016. "Supervised Learning for Detection of Duplicates in Genomic Sequence Databases." *pLoS ONE* 11, no. 8 (e0159644): 1-20.

Ciolino, J. D., H. L. Palac, A. Yang, M. Vaca, and H. M. Belli. 2019. "Ideal vs. Real: A Systematic Review on Handling Covariates in Randomized Controlled Trials." *BMC Medical Research Methodology* 19: 136. https://doi.org/10.1186/s12874-019-0787-8.

Cohen, S., R. Kamarck, and R. Mermelstein. 1983. "A Global Measure of Perceived Stress." *Journal of Health and Social Behavior* 24 (4): 385-396.

Cutrona, C., and D. Russell. 1987. "The Provisions of Social Relationships and Adaptation to Stress." *Advances in Personal Relationships,* 1.

Dadgar, M., and M. J. Trimble. 2015. "Labor Market Returns to Sub-Baccalaureate Credentials: How Much Does a Community College Degree or Certificate Pay?" *Educational Evaluation and Policy Analysis* 37 (4): 399-418. https://doi.org/10.1093/jeea/jvx028.

Deke, J., and H. Chiang. 2017. "The WWC Attrition Standard: Sensitivity to Assumption and Opportunities for Refining and Adapting to New Contexts." *Evaluation Review* 41: 130-154.

Deville, J. C., and C. E. Särndal. 1992. "Calibration Estimation in Survey Sampling." *Journal of the American Statistical Association* 87: 376-382.

Dundar, A., and D. Shapiro. 2016. *The National Student Clearinghouse as an Integral Part of the National Postsecondary Data Infrastructure.* Herndon, VA: National Student Clearinghouse Research Center. https://nscresearchcenter.org/wp-content/uploads/NSC-as-an-Integral-Part-of-the-National-Postsecondary-Data-Infrastructure.pdf

Dynarski, S. M., S. W. Hemelt, and J. M. Hyman. 2015. "The Missing Manual: Using National Student Clearinghouse Data to Track Postsecondary Outcomes." *Educational Evaluation and Policy Analysis* 37 (1s): 53S-79S.

Gardiner, K., and R. Juras. 2019. *PACE Cross-Program Implementation and Impact Study Findings.* OPRE Report 2019-32. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/sites/default/files/documents/opre/pace_cross_program_implementation_and_impact_study_findings_final.pdf.

Fein, D. 2012. *Career Pathways as a Framework for Program Design and Evaluation: A Working Paper from the Pathways for Advancing Careers and Education (PACE) Project*. OPRE Report 2012-30. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/opre/report/career-pathways-framework-program-design-and-evaluation-working-paper-pathways.

Fein, D., and S. Dastrup. Forthcoming. *Benefits that Last: Long-Term Findings for the Year Up Program.* Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Fein, D., S. Dastrup, and K. Burnett. 2021. *Still Bridging the Opportunity Divide for Low-Income Youth: Year Up's Longer-Term Impacts.* OPRE Report 2021-56. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/opre/report/still-bridging-opportunity-divide-low-income-youth-year-ups-longer-term-impacts.

Fein, D., and J. Hamadyk. 2018. *Bridging the Opportunity Divide for Low-Income Youth: Implementation and Early Impacts of the Year Up Program, Appendices*. OPRE Report 2018-65. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/opre/resource/bridging-the-opportunity-divide-for-low-income-youth-implementation-and-early-impacts-of-the-year-up-program.

Fein, D., D. Judkins, and L. Buron. 2021. *PACE Six-Year Follow-Up Analysis Plan.* OPRE Report 2021-29. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/opre/report/pace-six-year-follow-up-analysis-plan.

Folsom, R. E. 1991. "Exponential and Logistics Weight Adjustments for Sampling and Nonresponse Error Reduction." In *Proceedings of the American Statistical Association, Social Statistics Section*, 197-202. Alexandria, VA: American Statistical Association.

Folsom, R. E., and A. C. Singh. 2000. "The Generalized Exponential Model for Sampling Weight Calibration for Extreme Values, Nonresponse, and Post-Stratification." In *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 598-603. Alexandria, VA: American Statistical Association.

Folsom, R. E., and M. Witt. 1994. "Testing a New Attrition Nonresponse Adjustment Method for SIPP." In *Proceedings of the American Statistical Association, Section on Survey Research Methods,* 428-433. Alexandria, VA: American Statistical Association.

Goldrick-Rab, S., and K. Sorensen. 2010. "Unmarried Parents in College." *Future of Children* 20 (2): 179-203.

Hendra, R., and A. Hill. 2018. "Rethinking Response Rates: New Evidence of Little Relationship between Survey Response Rates and Nonresponse Bias." *Evaluation Review* https://doi.org/10.1177/0193841X18807719.

Holland, P. W. 1986. "Statistics and Causal Inference." Journal of the American Statistical Association 81 (396): 945–960. https://doi.org/10.1080/01621459.1986.10478354.

Hoven, M. R. 2012. "Investigating the Relationship between Perceived Social Support and Parent Self-Efficacy in Parents of Preschool-Aged Children." Master's Thesis. University of British Columbia. https://open.library.ubc.ca/soa/cIRcle/collections/ubctheses/24/items/1.0073240.

Judkins, D. 2019. "Covariate Selection in Small Randomized Studies." Presentation at the Joint Statistical Meetings, Denver, CO, July 2019. https://ww2.amstat.org/meetings/jsm/2019/onlineprogram/AbstractDetails.cfm?abstractid=307372.

Judkins, D., S. Dastrup, R. Juras, and M. Farrell. 2020. *The San Diego Workforce Partnership's Bridge to Employment in the Healthcare Industry Program: Appendices for Three-Year Impact Report.* OPRE Report 2020-105. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/opre/report/san-diego-workforce-partnerships-bridge-employment-healthcare-industry-program-three.

Judkins, D., D. Fein, and L. Buron. 2018. *Analysis Plan for the PACE Intermediate (Three-Year) Follow-Up Study*. OPRE Report 2018-95. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/opre/resource/analysis-plan-for-the-pace-intermediate-three-year-follow-up-study.

Judkins, D., D. Litwok, and K. Gardiner. 2020. *Pima Community College's Pathways to Healthcare Program: Appendices for Three-Year Impact Report.* OPRE Report 2020-43. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/opre/report/pima-community-colleges-pathways-healthcare-program-three-year-impact-report.

Judkins, D., D. Morganstein, P. Zador, A. Piesse, B. Barrett, and P. Mukhopadhyay. 2007. "Variable Selection and Raking in Propensity Scoring." *Statistics in Medicine* 26: 1022-1033.

Judkins, D. R., and K. E. Porter. 2016. "Robustness of Ordinary Least Squares in Randomized Clinical Trials." *Statistics in Medicine* 35 (11): 1763-1773.

Judkins, D., D. Walton, G. Durham, D. Litwok, and S. Dastrup. 2021. *Still Bridging the Opportunity Divide for Low-Income Youth: Year Up's Longer-Term Impacts, Technical Appendices.* OPRE Report 2021-56. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/opre/report/still-bridging-opportunity-divide-low-income-youth-year-ups-longer-term-impacts.

Judkins, D., D. Walton, D. Litwok, G. Durham, and S. Dastrup. 2021. *Valley Initiative for Development and Advancement (VIDA): Appendices for Three-Year Impact Report.* OPRE Report 2021-96. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/opre/report/valley-initiative-development-and-advancement-vida-three-year-impact-report.

Juras, R., and L. Buron. 2021. *Summary and Insights from the Ten PACE and HPOG 1.0 Job Training Evaluations.* OPRE Report 2021-155. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/opre/report/summary-and-insights-ten-pace-and-hpog-10-job-training-evaluations-three-year-cross.

Kessler, R. C., G. Andrews, D. Mrocek, B. Ustun, and H. U. Wittchen. 1998. "The World Health Organization Composite International Diagnostic Interview Short-Form (CIDI-SF)." *International Journal of Methods in Psychiatric Research* 7 (4): 171-185.

Koch, G. G., C. M. Tangen, J.-W. Jung, and I. A. Amara. 1998. "Issues for Covariance Analysis of Dichotomous and Ordered Categorical Data from Randomized Clinical Trials and Non-parametric Strategies for Addressing Them." *Statistics in Medicine* 17: 1863-1892.

Le, H., A. Casillas, S. Robbins, and R. Langley. 2005. "Motivational and Skills, Social, and Self-Management Predictors of College Outcomes: Constructing the Student Readiness Inventory." *Educational and Psychological Measurement* 65 (3): 482-508.

Lin, W. 2013. "Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique." *The Annals of Applied Statistics* 7: 295-318.

Litwok, D., D. Walton, and L. Peck. 2021. *Health Profession Opportunity Grants (HPOG) Impact Study's Six-Year Follow-Up Analysis Plan*. OPRE Report 2021-26, Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/opre/report/health-profession-opportunity-grants-hpog-impact-studys-six-year-follow-up-analysis.

Lumley, T., P. Diehr, S. Emerson, and L. Chen. 2002. "The Importance of the Normality Assumption in Large Public Health Data Sets." *Annual Review of Public Health* 23: 151-169.

Mosteller, F., and D. L. Wallace. 1963. "Inference in an Authorship Problem." *Journal of the American Statistical Association* 58, no. 302: 275-309.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel et al. 2011. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research* 12: 2825-2830. https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf.

Peterson, C. H., A. Casillas, and S. B. Robbins. 2006. "The Student Readiness Inventory and the Big Five: Examining Social Desirability and College Academic Performance." *Personality and Individual Difference* 41 (4): 663-673.

Raffo, J. 2020. "MATCHIT: Stata Module to Match Two Datasets Based on Similar Text Patterns." Statistical Software Components S457992. Boston College Department of Economics.

Rennie, J. D. M., L. Shih, J. Teevan, and D. R. Karger. 2003. "Tackling the Poor Assumptions of Naïve Bayes Text Classifiers." In *ICML '03 Proceedings of the Twentieth International Conference on Machine Learning*, edited by T. Fawcett and N. Mishra, 616-623. Palo Alto, CA: AAAI Press.

Research Triangle Institute. 2012. *SUDAAN Language Manual, Volumes 1 and 2, Release 11*. Research Triangle Park, NC: Author.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley.

Santor, D., and Coyne, J. 1997. "Shortening the CES-D to Improve Its Ability to Detect Cases of Depression. *Psychological Assessment* 9 (3): 233.

Stevens, A.H., M. Kurlaender, and M. Grosz. 2019. "Career technical education and labor market outcomes evidence from California community colleges." *Journal of Human Resources*, 54 (4): 986-1036.

Tukey, J. W. 1991. "Use of Many Covariates in Clinical Trials." *International Statistical Review* 59 (2): 123-137.

Zhang, H. 2004. "The Optimality of Naive Bayes." In *Proceedings of the Seventeenth International Florida Artificial Intelligence Research S*ociety Conference, edited by V. Barr and Z. Markov, 562-567. Palo Alto, CA: AAAI Press. https://www.aaai.org/Library/FLAIRS/flairs04contents.php.