Contract No.: AK-13690-04-30
MPR Reference No.: 6066-521

MATHEMATICA
Policy Research, Inc.

National Evaluation of the Trade Adjustment Assistance Program: Methodological Appendices Regarding the Baseline Survey (Covering Workers Eligible Under the 2002 Program)

Final Report

May 2010

Peter Z. Schochet

Submitted To:

U.S. Department of Labor
Employment and Training Administration
Room N-5637
200 Constitution Ave., NW
Washington, DC 20210

Project Officer:
Charlotte Schifferes

This report has been funded, either wholly or in part, with Federal funds from the U.S. Department of Labor (USDOL), Employment and Training Administration (ETA), Office of Policy Development and Research (OPDR) under Contract Number AK-13690-04-30. The report has been edited by OPDR. The contents of this publication do not necessarily reflect the views or policies of the Department of Labor, nor does mention of trade names, commercial products, or organizations imply endorsement of same by the U.S. Government.

Submitted By:

Mathematica Policy Research
(Subcontractor)
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005

Prime Contractor:
Social Policy Research Associates
1330 Broadway, Suite 1426
Oakland, CA 94612

Project Director:
Ron D'Amico

Principal Investigators:
Ron D'Amico
Peter Z. Schochet

CONTENTS

Methodological Appendices **Page**

INTRODUCTION ix

A THE SAMPLE DESIGN.....1

B THE BASELINE SURVEY.....29

C CONSTRUCTION OF SAMPLE WEIGHTS AND STANDARD ERRORS FOR BASELINE ANALYSES.....45

REFERENCES.....61

TABLES

Table		Page
A.1	STATE SELECTION PROBABILITIES FOR THE TAA EVALUATION	9
A.2	SELECTED STATES FOR THE TAA EVALUATION, BY REGION	11
A.3	UI COVERAGE RATES DURING THE TWO-YEAR POST-CERTIFICATION PERIOD FOR PETITIONS ASSOCIATED WITH THE SURVEY SAMPLE	20
A.4	COUNTS AND SHARES OF WORKERS IN THE SAMPLE UNIVERSE, BY STATE AND TAA PARTICIPATION STATUS	23
A.5	WORKER COUNTS IN THE SURVEY SAMPLE, BY STATE AND INITIAL TAA PARTICIPATION STATUS	27
B.1	DISTRIBUTION OF BASELINE SURVEY RELEASE DATES, BY TAA PARTICIPATION STATUS	32
B.2	RESPONSE RATES TO THE BASELINE SURVEY, BY TAA PARTICIPATION STATUS AND KEY SUBGROUP	36
B.3	DISTRIBUTION OF THE NUMBER OF MONTHS BETWEEN THE BASELINE SURVEY RELEASE DATE AND COMPLETION OF THE SURVEY, BY TAA PARTICIPATION STATUS	39
B.4	COMPARISON OF THE CHARACTERISTICS OF RESPONDENTS AND NONRESPONDENTS TO THE BASELINE SURVEY, BY TAA PARTICIPATION STATUS	41
C.1	DISTRIBUTION OF WEIGHTS, BY TAA PARTICIPATION STATUS	51
C.2	DISTRIBUTION OF THE NUMBER OF MONTHS BETWEEN THE UI CLAIM AND PETITION CERTIFICATION DATE	52

FIGURE

Figure **Page**

A.1 OVERVIEW OF STUDY DESIGN.....5

INTRODUCTION

In a series of appendices, this report provides an overview of methodological issues on the sample design and baseline survey for the Evaluation of the Trade Adjustment Assistance (TAA) Program. The methodological appendices complement the TAA evaluation report entitled, "The Characteristics of the Workers Eligible Under the 2002 TAA Program and Their Early Program Experiences." A more detailed discussion of these topics will be provided in a forthcoming TAA study implementation report.

The report contains the following three methodological appendices:

1. **"The Sample Design."** This appendix discusses the design that was used to obtain a nationally representative sample of states and eligible TAA workers within those states.
2. **"The Baseline Survey."** This appendix provides a discussion of the design of the baseline survey and survey response rates.
3. **"Construction of Sample Weights and Standard Errors for Baseline Analyses."** This appendix discusses the calculation of sample weights so that estimates based on baseline survey data can be generalized to the full study population. It also discusses the construction of standard errors of the estimates presented in the companion baseline report that account for design effects due to weighting and clustering.

METHODOLOGICAL APPENDIX A

THE SAMPLE DESIGN

A. INTRODUCTION

The key goal of the TAA evaluation is to use survey and administrative wage records data to obtain unbiased estimates of the impact of TAA on participants' employment-related outcomes. Impacts will be estimated for (1) the full sample, (2) subgroups of participants defined by their demographic characteristics, (3) subgroups of participants defined by their receipt of specific program services and benefits, and (4) subgroups of sites defined by key program features.

The ideal design—random assignment—was not feasible for the TAA evaluation, because TAA services cannot be denied to eligible workers under current program rules (so that it would not be possible to construct a control group). Furthermore, it was not feasible to randomly assign participants to different service groups, because TAA services are voluntary and are tailored to meet the needs of individual clients. Consequently, the evaluation is employing a comparison group (propensity score matching) design to obtain estimated impacts.

The sample design for the TAA impact evaluation has been structured to achieve several critical analysis objectives. First, it was structured to produce a sample that is representative of the national population of workers who are eligible for and receive TAA services and benefits. Second, it was structured to produce a sample that is representative of the national population of TAA-eligible nonparticipants to estimate program take-up rates and reasons for program participation and nonparticipation (a key topic that is discussed in the companion baseline report). Third, it was structured to generate a comparison sample of dislocated workers who are as similar as possible to workers in the TAA samples, except for the offer of TAA services. This comparison sample will be used to assess what the outcomes of treatment group members would have been in the absence of the TAA program. Finally, the sample design was structured to provide sufficient statistical precision for estimating policy-relevant program impacts.

This methodological appendix discusses key elements of the sample design that pertain to the companion descriptive baseline report that presents descriptive information on the characteristics of eligible TAA workers and their initial experiences with the TAA program. This analysis was conducted using the *TAA (treatment) sample* and *baseline survey data*, and thus, the focus of this appendix is on the design that was employed to obtain the baseline survey sample.

This appendix is in five sections. Section B provides a summary of the design. Section C discusses the selection of states for the study, and Section D discusses key administrative state data that were used to obtain the study samples. Section E discusses the sample frame for the study, and the final section discusses the selection of the sample that was released for baseline interviewing. A forthcoming report on study implementation will provide more details on these topics, and will also discuss the selection of the matched comparison group and the selection of the sample for administrative wage records data collection.

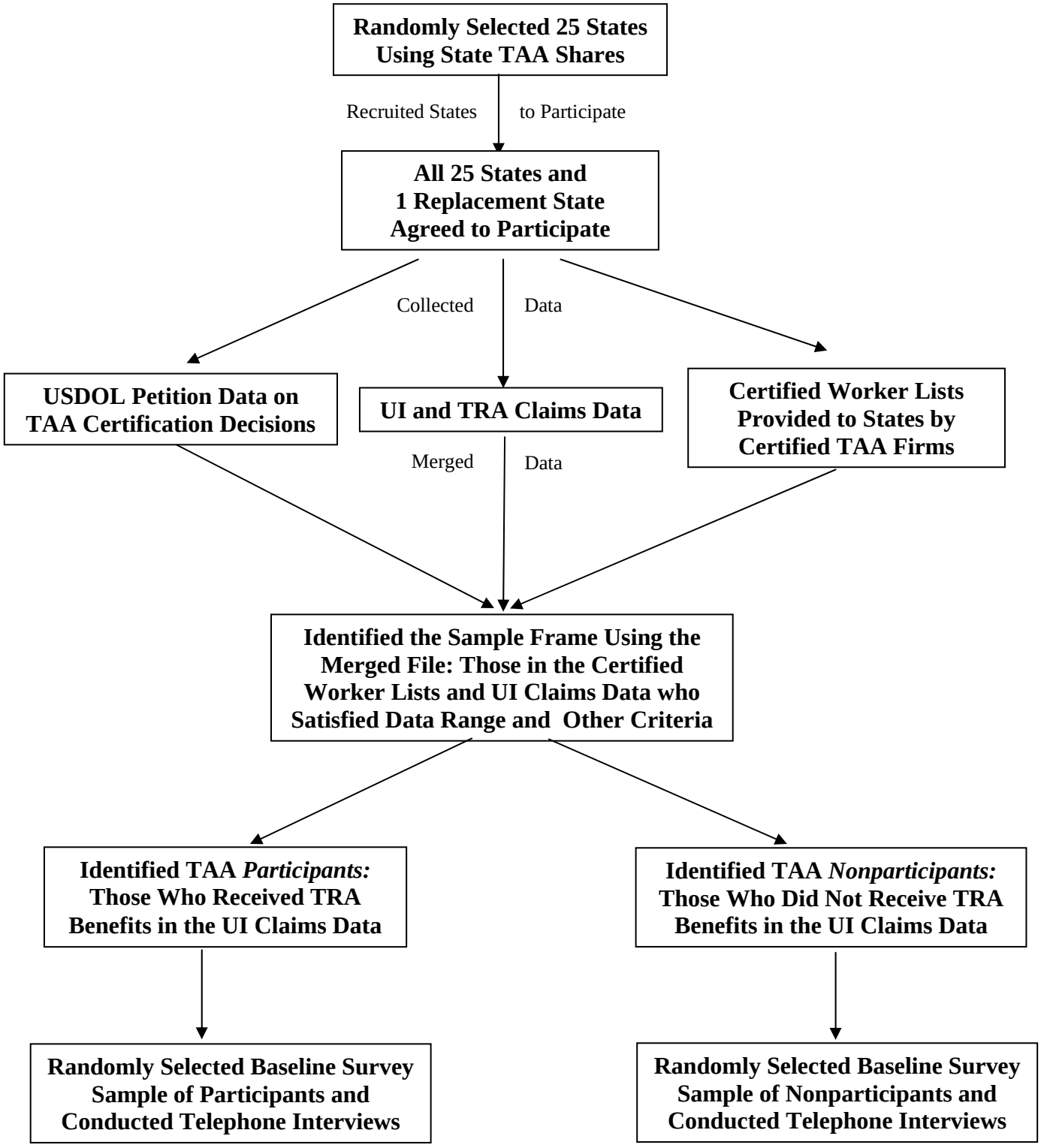
B. OVERVIEW OF DESIGN

Figure A.1 displays a flow chart of our design for selecting a nationally representative sample of eligible TAA workers for the baseline survey. The main steps were as follows:

1. 25 states were randomly selected in geographic strata with probabilities proportional to the expected number of TAA participants in the state.
2. Recruitment efforts resulted in all 25 states agreeing to participate in the study, along with 1 additional state that was recruited as a replacement state due to the initial reluctance of some states to participate in the study.
3. Data from two sources were collected from the 26 study states that were used to define the sampling frame for the study. First, data on TAA-eligible workers were collected from lists of trade-affected workers provided to state agencies by firms who were certified for TAA. Second, UI claims data were collected from each state.
4. TAA petition data were collected from the U.S. Department of Labor (USDOL). These data contain historical information on certification decisions and dates for all petitions submitted to USDOL. These data were used to identify workers from firms whose petitions were certified within our sampling frame window.

FIGURE A.1

OVERVIEW OF STUDY DESIGN



5. The certified-worker lists and UI claims data were merged by SSN (or name and zip code if SSN was not available), and this file was then merged to the USDOL petition file by petition number (which was available in the certified-worker lists).

6. The sample frame was identified using the merged file and includes those in both the certified-worker lists and the UI claims data who satisfied date range and other sample frame criteria. Thus, the sample frame for the "*certified-worker sample*" contains workers who were laid off from TAA-certified firms during the period covered by certification, and who subsequently received UI benefits.

7. Within each state, the sample frame was divided into "*TAA participants*" (those who received Trade Readjustment Allowance [TRA] benefits according to the UI claims records) and "*TAA nonparticipants*" (those who did not receive TRA benefits according to the UI claims records).

8. Within each state, the "*baseline survey sample*" was obtained by randomly selecting separate subsamples of TAA participants and TAA nonparticipants using stratified sampling techniques. Twice as many TAA participants than nonparticipants were selected for baseline interviewing. The contact information in the UI claims data was used to locate sample members for telephone interviews. A single round of follow-up surveys will be conducted with participants only, and UI wage records will be collected for both participants and nonparticipants.

The remainder of this appendix discusses these steps in more detail.

C. SELECTION OF STATES

This section first discusses the state selection design, and then briefly discusses the state recruitment process (which will be discussed in more detail in the forthcoming report on study implementation).

1. State Selection Design

Our design called for selecting a random subset of states rather than from all states nationwide, for two reasons: (1) the TAA caseload is relatively concentrated, and (2) sample selection and data acquisition costs would have increased significantly with the number of states selected. Although a clustered sample of states results in a slight loss in the precision of study estimates (but no bias), the savings in resources and reduced administrative complexity provided by sampling states more than offset this loss.

To select the states for the evaluation, we obtained from USDOL petition data on all TAA and NAFTA industry certifications from fiscal year (FY) 1999 through FY 2006. These petition data provided a sample frame from which to select the states, because each petition contains information on the estimated number of trade-affected workers (that is, those who are likely to lose their jobs in the period covered by the certification). The petition data contain information on more than 14,200 certified firms, covering nearly 1.5 million dislocated workers.

Although the study included workers from firms whose petitions were certified during the one-year period from November 1, 2005 to October 31, 2006, we collected petition data from multiple years to examine the extent to which state shares of the eligible TAA population changed over time. This analysis was important for several reasons. First, we wanted to set state sampling probabilities that were based on "typical" state shares to "smooth out" unusually high or low state TAA activity in a given year. For example, we did not want to assign a low sampling probability to a state that had an unusually low TAA share in FY 2006, but that had much higher shares in FY 1999 to FY 2005. Second, the information in the petitions on the estimated number of trade-affected workers is known to be somewhat noisy. Thus, using historic petition data could help remove this noise, and yield more accurate estimates of actual state shares during the period covered by the study.

The trend analysis revealed that state shares were relatively constant over time; that is, states with relatively high TAA activity in one year tended to have relatively high TAA activity in other years. For instance, from FY 2003 to FY 2006, the correlation between state shares in any two years was about .85, and similarly for the correlations between state share rankings. In addition, there was little change over time in the 15 or 20 states with the largest TAA worker shares.

Given these analysis findings, we randomly selected states using the *average* of the state shares in FY 2005 and FY 2006. Table A.1 displays these state shares (Column 3) and state selection probabilities (Column 4) assuming sampling with replacement. The figures pertain to the 50 states, the District of Columbia, and Puerto Rico. The state selection probabilities sum to 25, the number of states originally selected for the study. The table also displays selection probabilities that sum to 26 (Column 5), because the final study sample included one additional (replacement) state that was approached in the recruitment phase of the study and that agreed to participate in the evaluation (see below). For simplicity, this 26-state design was "assumed" for calculating sample weights (see Appendix C). The data are ordered by state, according to their shares of the TAA population, from largest to smallest.

Using Table A.1, we randomly selected 25 original states with probabilities proportional to the state shares shown in Column 3. Thirteen states (North Carolina, California, Pennsylvania, Michigan, South Carolina, Georgia, Tennessee, Ohio, Illinois, Indiana, Texas, New York, and Alabama) were chosen with certainty.[1] Four additional states (Kentucky, Virginia, Wisconsin, and Missouri) were also chosen with certainty, because after removing the initial thirteen certainty states, the probability of selecting these four states was .96, .96, .88, and .87, respectively. State selection occurred in late 2006.

The remaining 8 noncertainty states were randomly sampled from the universe of 35 noncertainty states, with the probabilities shown in Column 4 of Table A.1. We selected the noncertainty states by ordering them by the six USDOL regions and using a systematic sampling approach; this ensured that the sample of states would be dispersed geographically. Geographic

[1] The nine states with initial weights greater than 1 were chosen with certainty, because these states had more than 1/25 of the total weight. After removing these nine states, we also chose four additional states with certainty, because they had more than 1/16 (1 ÷ [25–9]) of the remaining total weight.

TABLE A.1

STATE SELECTION PROBABILITIES FOR THE TAA EVALUATION

State	USDOL Region	Average Annual Share of Trade-Affected Workers in Certified Firms in FY 2005 and FY 2006 (Percentages)[a]	State Selection Probability Under a 25-State Design	State Selection Probability Under a 26-State Design
North Carolina	3	9.7812	1.0000	1.0000
California	6	9.5307	1.0000	1.0000
Pennsylvania	2	5.7822	1.0000	1.0000
Michigan	5	5.6956	1.0000	1.0000
South Carolina	3	4.8528	1.0000	1.0000
Georgia	3	4.7894	1.0000	1.0000
Tennessee	3	4.5840	1.0000	1.0000
Ohio	5	4.4514	1.0000	1.0000
Illinois	5	4.2700	1.0000	1.0000
Indiana	5	3.9740	1.0000	1.0000
Texas	4	3.6127	1.0000	1.0000
New York	1	3.5500	1.0000	1.0000
Alabama	3	3.0492	1.0000	1.0000
Kentucky	3	2.5598	1.0000	1.0000
Virginia	2	2.5555	1.0000	1.0000
Wisconsin	5	2.3617	1.0000	1.0000
Missouri	5	2.3319	1.0000	1.0000
Massachusetts	1	1.9201	0.6898	0.7760
Arkansas	4	1.8641	0.6697	0.7534
New Jersey	1	1.4914	0.5358	0.6028
Oklahoma	4	1.4737	0.5294	0.5956
Mississippi	3	1.2177	0.4375	0.4922
Minnesota	5	1.1652	0.4186	0.4709
Colorado	4	1.1638	0.4181	0.4704
Iowa	5	1.0916	0.3922	0.4412
Oregon	6	1.0808	0.3883	0.4368
Florida	3	1.0023	0.3601	0.4051
New Hampshire	1	0.9446	0.3393	0.3818
Maryland	2	0.8953	0.3216	0.3619
West Virginia	2	0.8616	0.3095	0.3482
Rhode Island	1	0.8310	0.2985	0.3359
Washington	6	0.8246	0.2963	0.3333
Connecticut	1	0.7194	0.2585	0.2908
Arizona	6	0.5757	0.2068	0.2327
Maine	1	0.5018	0.1803	0.2028
Vermont	1	0.3782	0.1359	0.1528
Kansas	5	0.3318	0.1192	0.1341
Idaho	6	0.2475	0.0889	0.1000
Utah	4	0.2276	0.0818	0.0920
Arkansas	4	0.2034	0.0731	0.0822
Nevada	6	0.1940	0.0697	0.0784
Nebraska	5	0.1828	0.0657	0.0739
Louisiana	4	0.1784	0.0641	0.0721
Delaware	2	0.1663	0.0597	0.0672
South Dakota	4	0.1587	0.0570	0.0641
Montana	4	0.1200	0.0431	0.0485
Puerto Rico	1	0.0973	0.0350	0.0393
Hawaii	6	0.0634	0.0228	0.0256
New Mexico	4	0.0515	0.0185	0.0208
North Dakota	4	0.0429	0.0154	0.0173
Wyoming	4	0.0000	0.0000	0.0000
District of Columbia	2	0.0000	0.0000	0.0000
Total		**100.0000**	**25.0000**	**26.0000**

Source: DOL Petition Data on all Industry Certifications in FY 2005 and FY 2006.

[a]Figures pertain to the estimated number of trade-affected workers that are denoted in each petition.

stratification was a useful way of ensuring that the sample of states would represent the full range of TAA programs and participants, because states within a geographic area tend to have similar industries, workers, and labor markets. The selected noncertainty states were as follows: *Region 1*: New Hampshire, New Jersey, and Rhode Island; *Region 3:* Florida; *Region 4:* Arkansas and Colorado; *Region 5:* Minnesota; and *Region 6:* Washington.

After selecting the 25-state sample, we also randomly selected "replacement" states in the event that "primary" states refused to participate in the study. We sequentially randomly selected replacement states within each region using the sampling techniques discussed above. The plan was to contact replacement states in a region (moving down the ordered list) if we could not solicit the cooperation of the primary states in that region. This process yielded the sample of primary and ordered replacement states shown in Table A.2.

As discussed further below, based on *actual* data collected from the 26 states, we estimate that the 17 certainty states contain about 79 percent of all TAA-eligible workers in the sample frame for the study. The corresponding figure is 10 percent for the 9 noncertainty states (including the replacement state Maryland). Consequently, the total sample of 26 certainty and noncertainty states contains nearly 90 percent of all workers in the sample frame.

2. State Recruitment

State recruitment started in early 2007 and involved contacting senior regional and state Labor Department officials, and state TAA coordinators and administrators. The study team conducted initial telephone calls with regional and state staff, explaining the nature and importance of the study and its data requirements. Study materials were subsequently sent to the states describing the evaluation and data requests in more detail.

All 25 selected states eventually agreed to participate in the study. However, it typically took many months and considerable involvement by USDOL and evaluation staff to solicit the

TABLE A.2

SELECTED STATES FOR THE TAA EVALUATION, BY REGION

25-State Sample	Replacement States (in Order of Selection)
Region 1	Connecticut, Massachusetts, Vermont, Maine, Puerto Rico
New Hampshire	
New Jersey	
New York[a]	
Rhode Island	
Region 2	Maryland, Delaware, West Virginia, Washington D.C.
Pennsylvania[a]	
Virginia[a]	
Region 3	Florida, Mississippi
Alabama[a]	
Georgia[a]	
Kentucky[a]	
North Carolina[a]	
South Carolina[a]	
Tennessee[a]	
Region 4	Utah, Oklahoma, Montana, Louisiana, South Dakota, North Dakota, New Mexico, Wyoming
Texas[a]	
Arkansas	
Colorado	
Region 5	Iowa, Nebraska, Kansas
Illinois[a]	
Indiana	
Michigan[a]	
Minnesota	
Missouri[a]	
Ohio[a]	
Wisconsin[a]	
Region 6	Arizona, Oregon, Idaho, Arkansas, Hawaii, Nevada
California[a]	
Washington	

[a]Denotes a certainty state.

cooperation of states, to obtain Memorandums of Understanding (MOUs) with the states, and to obtain the requested data from the states. The primary reasons why states were initially reluctant to participate in the evaluation were (1) they did not have enough programming resources to provide the considerable amounts of longitudinal administrative data that were requested for the study, and (2) legal issues surrounding releasing confidential data. These issues were resolved through negotiations between state and USDOL lawyers, and by amending our data requests to make them as simple and non-burdensome as possible. For example, the study team negotiated that the states could provide "data dumps" of Unemployment Insurance (UI) claims data, rather than study-specific data extracts whose construction would have required more programmer resources.

USDOL, SPR, and Mathematica were typically signatories to the state MOUs. The MOUs specified the data to be provided by pertinent state agencies, compensation and costs for the data, the way in which the data are to be used for the study, and data security and confidentiality provisions.

The states provided the first round of data throughout 2008. These data included the certified-worker lists and the UI claims data, which were needed to identify the sample frame for the certified-worker sample. Complete data were provided by 4 states in the first quarter of 2008, 10 states in the second quarter of 2008, 6 states in the third quarter of 2008, and 6 states in the fourth quarter of 2008. All states sent the requested data to the study contractors, except California, where study programmers selected the study samples on site and copied, to CDs, pertinent information for these samples only. None of the state MOUs specified a maximum study sample size, except California.

Finally, during the recruitment phase, there was considerable uncertainty as to which of the 25 selected states would ultimately participate in the evaluation. Once we realized the time it was

going to take to obtain final responses from the 25 states and the protracted recruitment and negotiation process, we contacted several replacement states (using the ordered list shown in Table A.2). Replacement states were first contacted in regions where recruitment efforts for the primary states were progressing slowly. During this process, Maryland (the first replacement state in Region 2) agreed to participate in the study and an MOU was established. Therefore, USDOL decided to include Maryland in the study. Thus, the final sample has 26 states rather than 25. As shown in Table A.1, the sampling probabilities are similar for a 25-state or 26-state design. Thus, for simplicity, we "assume" for the analysis that 26 states were originally randomly selected for the study.

D. SUMMARY OF KEY DATA PROVIDED BY THE STUDY STATES

The sample frame for the *certified-worker sample* contains workers who were laid off from TAA-certified firms during the period covered by certification, and who subsequently received a first UI payment. This section discusses the two primary sources of administrative state data that were used to define this sample frame: (1) certified-worker lists and (2) UI claims data.

1. Certified-Worker Lists

The sample frame for the certified-worker sample was obtained using all potentially TAA-eligible workers in lists that TAA-certified firms provided to the 26 states included in the evaluation. These lists are available (and include the workers' contact information) because, under the 1988 legislative changes to the TAA program, state agencies became required (1) to identify potentially eligible workers by obtaining lists of workers who were separated or partially separated from trade-affected firms during the period covered by certification, and (2) to notify each potentially eligible worker in writing. Workers covered by a certification include those laid off between one year prior to the petition *filing* date and two years after the petition *certification*

date. This typically translates into a three to three-and-one-half-year layoff period, because it often takes several months for USDOL to make certification determinations.

As discussed further below, our sample includes workers from firms whose petitions were certified during the one-year period from November 1, 2005 to October 31, 2006. Thus, we requested that states provide us with lists that covered this time period, although it was easier for most states to provide lists covering a longer period. The certified lists were typically provided by states in EXCEL spreadsheets or hardcopy form.

The certified-worker lists usually contain information on the SSN, name, and address of each worker. These data items were critical for matching workers to the UI claims data to select the certified-worker samples (as discussed below). For each worker in the certified-worker lists, there is also information on the *TAA petition number* for the worker's firm. This petition number was used to merge the certified-worker lists to a petition file provided by USDOL that contains historic information on each petition submitted to USDOL. This petition file contains certification decisions and dates, which were critical for defining the sample frame for the study.

As discussed in detail in the forthcoming report on study implementation, the certified-worker lists appear to be somewhat comprehensive, and thus, constitute a reasonable sampling frame for the TAA study. The evidence suggests that most certified petitions in the USDOL petition file were covered in the certified-worker lists, many workers listed in the certified-worker lists were matched to a record in the UI claims files, and worker counts in the certified-worker lists were somewhat consistent with TRA beneficiary counts in the UI claims data. However, it is important to emphasize that our results generalize formally only to those workers listed on the certified-worker lists (and who could be matched to records in the UI claims data), and not necessarily to all trade-affected workers.

2. UI Claims Data

UI claims data were critical for the evaluation because they were used to:

- Define the sample frame for the study by restricting those in the certified-worker lists to those who received UI benefits.
- Provide information on the receipt of TRA benefits that was needed to classify members of the certified-worker sample as TAA participants or TAA nonparticipants (see below).
- Identify comparison group members, who consist of UI recipients who were matched to treatment group members based on information contained in the UI claims files (not discussed here)
- Provide contact information (name, address, telephone number, and SSN) that was needed to locate TAA (and comparison group) members for baseline surveys.

Most states provided data dumps of all workers who received a first UI payment of any type from the first quarter of 2004 to the most recent quarter that UI records were available when the data were extracted. The data coverage period differs somewhat across states; the most recent period covered by the UI data was the first quarter of 2008 for three states, the fourth quarter of 2007 for 10 states, one of the first three quarters of 2007 for 9 states, and the fourth quarter of 2006 for 4 states. The UI files were typically quite large; the number of records ranged from about 68,000 in New Hampshire to 2.1 million in Pennsylvania.

The UI claims data for each state contain the following information:

- ***Identifying information:*** SSN, name, address, and telephone number
- ***Demographic information:*** Gender, date of birth, and race/ethnicity
- ***Job characteristics:*** Base-period earning and industry of main base-period employer.
- ***UI claim and benefit data:*** Benefit year begin date; date of UI or TRA first payment; date of UI or TRA last payment; UI claim type (regular UI, emergency UI, TRA, etc.); UI and TRA maximum benefit amount; UI and TRA weekly benefit amount; and UI and TRA remaining claim balance.

About half the states also provided worker profiling information (such as profiling scores), and a few states also provided additional information, such as weeks worked on the job or claimant's education level.

Finally, for purposes of matching TAA participants to comparison group members and for creating subgroups for analysis, we merged, by state, county, and year (if relevant), the following local area characteristics into the UI claims records:

- ***The annual unemployment rate in 2000 to 2006*** using data from the U.S. Bureau of Labor Statistics (BLS).
- ***The poverty rate in 2004*** using data from the Area Resource File (ARF).
- ***The percentage of workers in manufacturing in 2000 and 2005*** using ARF data.
- ***The average earnings per job in 2005*** using data from the Inter-University Consortium for Political and Social Research (ISPSR).
- ***The percentage population growth between July 1,/2000 and July 1, 2005*** using ICPSR data.
- ***The U.S. Department of Agriculture, Economic Research Service (ERS) 2003 Rural-Urban Continuum Code*** using ICPSR data. These codes form a classification scheme that distinguishes metropolitan counties by the population size of their metropolitan area, and nonmetropolitan counties by degree of urbanization and adjacency to a metropolitan area or areas. There are nine such codes that range from a metropolitan area with a population of 1 million or more to rural areas that are not adjacent to a metropolitan area.
- ***Local area unemployment statistics (LAUS) area type indicators in 2007*** using BLS data. These indicators pertain to labor market areas that are economically integrated geographic areas within which individuals can reside and find employment within a reasonable distance or can readily change employment without changing their place of residence. Labor market areas are metropolitan areas, micropolitan areas, or small labor market areas, and exhaust the geography of the U.S. These area definitions are often used to allocate Federal program funds to states and local areas.

3. Merging Files

To obtain the sample frame for the certified-worker sample and contact information for baseline interviewing, it was necessary to merge the certified-worker lists, the USDOL petition

file, and the UI claims data. The certified-worker lists were first merged to the USDOL petition file (which contains certification decisions and dates) using the TAA petition number. The resulting file was merged to the UI claims data using SSNs, or in cases where SSNs were not available, using name and zip code (as discussed in more detail in the forthcoming report on study implementation).

It some instances, a worker matched to more than one record in the UI claims data. In these cases, we defined rules to identify the UI record that was most likely to be associated with the pertinent TAA petition (for example, by selecting the UI record with an associated TRA claim). Workers were excluded from the study who (1) did not match to a UI record or (2) matched to multiple UI records, but for whom we could not confidently identify the "correct" record.

E. THE SAMPLE FRAME FOR THE CERTIFIED-WORKER SAMPLE

This section first discusses the definition of the sample frame for the certified-worker sample, and then discusses our approach for separating the sample frame into "TAA participants" and "TAA nonparticipants." The final section presents counts of workers in the sample universe, by state and TAA participation status.

1. Defining the Sample Frame

The sample frame for the certified-worker sample was obtained using the merged file discussed above. The sample frame includes the following workers:

- ***Workers in the certified-worker lists who were laid off from firms that became certified for TAA between November 1, 2005 and October 31, 2006.*** As discussed, these workers were identified using the USDOL petition file. Importantly, even though states furnished data at different times, the petition certification period for the study was the *same* for all states. We specified a one-year window to account for potential seasonal layoff patterns.
- ***Those whose UI benefit year started in the approximately three-year period covered by their firms' TAA certification.*** The study included only UI recipients,

because few UI nonrecipients are eligible to receive TAA benefits. Furthermore, the comparison group sample was selected from UI recipients, so that UI claims records data were needed for matching purposes. Finally, the UI data provided contact information for the baseline survey.

Workers covered by a certification include those laid off between one year prior to the petition *filing* date and two years after the petition *certification* date. It typically takes USDOL about two months to make certification determinations. Thus, the sample frame for the certified-worker sample consists of workers whose UI benefit year started between September 1, 2004 and October 31, 2008.

- ***Workers between the ages of 16 and 80.*** Worker ages were calculated using UI data on birth dates and claim dates. A small number of workers had calculated ages that were outside the 16-to-80 range, and we suspect that some of these were due to data errors. Age was a critical variable for matching TAA sample members to comparison group members, for checking the identity of sample members at the start of the telephone survey, and for screening survey respondents for age-related survey questions (for example, questions on Alternative Trade Adjustment Assistance [ATAA] were asked only of those 50 or older). Thus, we excluded those outside the 16-to-80 age range and those with missing birth date.

- ***Workers who received regular UI benefits.*** UI records associated with special UI programs (such as emergency unemployment compensation, disaster unemployment assistance, and state and federal extended unemployment benefits) were excluded for the study. This is because these programs are atypical and could influence the types and amount of TAA services that are received by trade-affected workers. For instance, benefits from these special UI programs are typically paid *before* TRA payments are paid, which could influence TAA training decisions. Furthermore, these special UI benefits would be received by *both* treatment and comparison group members, which could result in smaller differences between the UI benefits received by the two research groups, and hence, smaller TAA impacts on training and employment-related outcomes. Less than 1 percent of all records had these claim types.

- ***Workers with nonmissing values for key data items.*** A very small number of cases were excluded who had missing or invalid data values for gender, base wages, the UI benefit year begin and first payment dates, the maximum benefit amount, the UI claim type, and zip code. Finally, for survey purposes, we excluded a small number of cases who did not have a telephone number in the UI claims data.

A crucial design decision was the time period over which to define the certified-worker universe. As discussed, workers covered by a certification include those laid off between one year prior to the petition filing date and two years *after* the petition certification date. This two-

year post-certification coverage period presented a challenge for the study design, because the UI data for some states do not extend beyond 2006. Thus, to ensure that our design would cover the full two-year post-certification coverage period for all states, we would have needed to select firms that were certified in 2004 or earlier.

We did not adopt this design, however, for several reasons. First, this design could have led to serious survey recall error, because sample members with a certification date in January 2004, for example, could have experienced their layoff as early as February 2003, which is more than *five* years before baseline interviewing started. Second, to rigorously evaluate the TAA program after the 2002 Act, a design goal was to conduct the evaluation after key provisions of the Act (such as the Health Care Tax Credit [HCTC] and ATAA programs) had been fully implemented for some time.

Instead, we selected a one-year petition certification period between November 1, 2005 and October 31, 2006, which results in a sample frame of TAA-eligible workers who received UI benefits between September 1, 2004 and October 31, 2008. Under this design, the UI data cover all workers who were laid off before the petition filing date and most workers during the two-year period after the filing date (because the UI data for most states cover the 2004 to 2007 period). As shown in Table A.3, the UI data cover 17 months of the 24-month post-certification period for the average petition in our sample. Furthermore, the UI data cover at least half of the post-certification period for three-quarters of the petitions (Table A.3). UI coverage rates, however, differ somewhat across states due to differences in the dates that the states extracted the data. Average state post-certification coverage rates range from 7 to 22 months, but nearly three-quarters of states have coverage rates that are longer than 15 months (not shown).

TABLE A.3

UI COVERAGE RATES DURING THE TWO-YEAR POST-CERTIFICATION PERIOD FOR PETITIONS ASSOCIATED WITH THE SURVEY SAMPLE

Number of Post-Certification Months Covered by the UI Claims Data	Percentage of Certified Petitions
9 to 12	12
13 to 16	18
17 to 20	28
21 to 23	17
24	14
(Mean Number of Months)	(17)
Number of Petitions	**524**

Source: UI claims data and USDOL petition files from the 26 study states.

Importantly, we found that only about 10 percent of workers started their UI spell more than 12 months after their firm's petition certification date. Stated differently, about 90 percent of trade-affected workers filed for UI either before or within 12 months after their firm became certified for TAA. These figures were computed by comparing UI claim and petition certification dates for those workers whose associated certification window was fully covered by the UI data.

These empirical analyses suggest that our sample is largely representative of trade-affected workers in our certified-worker universe, although not necessarily for workers who filed for UI benefits many months after the certification date. Furthermore, the representativeness of the sample is also somewhat uneven across states. To address these issues, we constructed weights to adjust for these forms of underrepresentation (see Appendix C).

2. The Sample Frame for TAA Participants and TAA Nonparticipants

The certified-worker lists contain information for TAA-eligible workers who received TAA services and for those who did not. Our evaluation focuses on both groups of workers, but the greater share of study resources are targeted to the participants.

The main purpose of the nonparticipant group for the study is to examine reasons for program nonparticipation and other (non-TAA) training-related services received by these workers (as discussed in the companion baseline report). In addition, the TAA program might have an effect on the earnings of these workers, because of TAA provisions that could increase their participation in other training programs. Consequently, we selected a comparison group for them and will include them in the sample for whom we will collect administrative UI wage records.

However, we expect *larger* program impacts for TAA participants than for nonparticipants. Furthermore, an important component of the evaluation is to describe fully the TAA experiences of program participants and to estimate impacts for participant subgroups defined by the receipt of specific TAA services. Thus, more survey resources are targeted to the participant group. In particular, we conducted *twice* as many baseline surveys with participants than nonparticipants, and will conduct 24-month follow-up interviews with participants only.

Because of this design, in order to select the baseline survey sample, we first needed to identify program participants and nonparticipants. We did this using UI records data information on *TRA benefit receipt*. TRA benefit receipt is likely to be a very good proxy for TAA program participation, because, as discussed in the companion baseline report, most of those who reported receiving some TAA services received TRA benefits.

Thus, we initially defined TAA participants and nonparticipants as follows:

1. ***Participants were defined as those who received TRA benefits according to the UI claims data.*** Most of the sample had a sufficient follow-up period to accurately determine whether they were TRA recipients after they exhausted their UI claims. This is because the sample was obtained from workers in firms that become certified for TAA between November 2005 and October 2006 (and, hence, workers laid off starting in September 2004), and the UI claims data from the states typically cover claims filed through 2007.

2. ***Nonparticipants were defined as those who had not received TRA benefits according to the UI claims data.*** This group includes those with a relatively short follow-up period between job loss and the most recent period covered by their states' UI data, as well as those with a longer follow-up period who never received TRA benefits.

The "nonparticipant" sample also contains some workers who ultimately received TAA services and benefits, as well as those who already received TAA training services but not TRA. We anticipated that about 20 percent of these initially-defined nonparticipants would actually be TAA participants. However, we did not know the identity of these "switchers" prior to the baseline survey. Thus, to adjust for these switchers, we released for baseline surveys proportionately more nonparticipants than participants to achieve our target sample sizes.

Using baseline survey data, the actual worker switching rate was 29 percent. Switchers were identified as those who reported in the baseline survey as having received any core TAA services: TRA, TAA-funded training, health coverage through the HCTC, and, for workers over age 50, wage subsidies through ATAA. We will update the TAA participant and nonparticipant designations before the 24-month follow-up survey using more recent TRA data, and if available, TAA program records.

3. Worker Counts in the Sample Universe

There were 49,531 workers in the certified-worker sample universe in the 26 study states (Table A.4). This figure includes 16,344 TAA participants and 33,187 TAA nonparticipants (based on initial TAA participation designations).

TABLE A.4

COUNTS AND SHARES OF WORKERS IN THE SAMPLE UNIVERSE, BY STATE AND TAA PARTICIPATION STATUS

	Worker Universe Counts			Share of All Workers in the Entire Universe (Percentage)	
Study State	TAA Participants	TAA Nonparticipants	Total	Using Certified-Worker Lists	Original Estimates
North Carolina	3,161	4,233	7,394	13.46	9.78
Pennsylvania	1,382	3,130	4,512	8.22	5.78
Georgia	1,409	2,502	3,911	7.12	4.79
Wisconsin	117	3,706	3,823	6.96	2.36
Ohio	701	3,082	3,783	6.89	4.45
California	477	3,016	3,493	6.36	9.53
Tennessee	1,192	1,276	2,468	4.49	4.58
Illinois	1,022	1,403	2,425	4.42	4.27
Arkansas	763	1,399	2,162	3.94	1.86
Alabama	1,003	1,000	2,003	3.65	3.05
Virginia	823	1,180	2,003	3.65	2.56
Michigan	320	1,269	1,589	2.89	5.70
New York	558	921	1,479	2.69	3.55
Indiana	878	350	1,228	2.24	3.97
South Carolina	431	638	1,069	1.95	4.85
Texas	347	698	1,045	1.90	3.61
Missouri	206	745	951	1.73	2.33
Kentucky	333	557	890	1.62	2.56
Colorado	178	574	752	1.37	1.16
New Jersey	373	344	717	1.31	1.49
Maryland	152	332	484	0.88	0.90
Florida	100	255	355	0.65	1.00
New Hampshire	80	248	328	0.60	0.94
Washington	132	153	285	0.52	0.82
Rhode Island	138	77	215	0.39	0.83
Minnesota	68	99	167	0.30	1.17
Total in the 26 Study States	**16,344**	**33,187**	**49,531**	**91.35**	**89.62**
Estimated Total in the Universe	**17,892**	**37,030**	**54,922**		

Source: Certified Worker Lists Provided by the 26 Study States.

These counts translate into a sample universe of 54,922 workers across the 50 states, the District of Columbia, and Puerto Rico (Table A.4). This figure was estimated using the following formula:

$$(1)\quad Total\ Number\ of\ Workers\ in\ Universe = \sum_{s=1}^{26} \frac{Total_s}{Prob_s},$$

where $Total_s$ is the worker count in state s and $Prob_s = 26 * Share_s$, where $Share_s$ is the estimated share of workers in state s that was used for state selection (see the final column in Table A.4 that reproduces state shares from the 26-state design shown in Table A.1 above). This universe includes 17,892 participants and 37,030 nonparticipants.

These estimates suggest that the 26 study states contain 90.2 percent of all workers in the certified-worker sample universe (49,531 workers in the 26 study states divided by the estimated 54,922 workers in the sample universe). Similarly, the sample contains 91.3 percent of all participants and 89.6 percent of all nonparticipants in the study universe.

As discussed, some initial nonparticipants reported receiving TAA services in the baseline survey and were reclassified as participants; the median state switching rate was 25 percent, but the rate ranged from 0 to 76 percent across the 26 states. After accounting for these switchers, we estimate that the sample universe contains 26,889 participants and 28,033 nonparticipants.

Finally, we find that state worker shares using the *actual* data are similar to the *estimated* state worker shares that were used to sample the 26 states (see the last two columns of Table A.4). The correlation between the two shares is 0.79, and 17 of the 18 states with the largest shares using the actual data were defined as certainty states for sampling (Arkansas is the lone exception). These findings suggest that design effects due to state-level weighting are not large (see Appendix C).

F. SELECTION OF THE SURVEY SAMPLE

The baseline survey sample was randomly selected from all workers in the sample universe. Within each state, sampling was performed separately for participants and nonparticipants (using initial participation designations). In addition, we used systematic sampling methods, where workers were ordered by gender, local labor market area, race/ethnicity, and age to ensure a representative survey sample within key population strata.

Our design was structured to select state sample sizes of participants and nonparticipants to generate survey samples that were as close to self-weighting as possible. This design was adopted to maximize the precision of the study estimates for a given sample size of workers. To achieve this goal, we initially calculated participant and nonparticipant sample sizes in each of the selected states using the figures in Table A.1 and the following formula:

$$(2)\ n_s = f\frac{N_s}{p_s},$$

where n_s is the number of TAA-eligible workers selected in state *s*, N_s is the total estimated number of TAA-eligible workers in state *s*, and p_s is the probability that state *s* was selected. The term *f* is the national sampling fraction for the population being sampled. Thus, the formula in (2) set the initial sample in each state (n_s) so that the probability of selection was *f* for all program-eligible workers. The total probability that a worker was selected is the probability the state was chosen (p_s) times the probability that a person was chosen in the state (n_s/N_s).

The value of *f* was selected so that the state samples summed to 2,220 for TAA participants and 1,110 for TAA nonparticipants. These targets were selected so that baseline surveys could be completed with 1,770 participants and 885 nonparticipants, assuming an 80 percent survey response rate.

These initial sample sizes, however, were amended for a number of reasons:

1. ***State samples were released for interviewing in waves.*** As discussed, states provided data at different times throughout 2008, and thus, survey samples were released in several waves (see Appendix B). During this process, it was often difficult to anticipate which states would ultimately provide data and when. Thus, in calculating worker sample sizes for a particular state, it was necessary to make assumptions about the ultimate state sample size. These estimates increased from 18 states (for the earliest states) to 22 states, to 25 states, and finally to 26 states. These changes affected state sample sizes.

2. ***The survey response rate was lower than anticipated.*** Our initial worker sample sizes were selected assuming an 80 percent response rate to the baseline survey. However, to achieve our target number of completed interviews, we increased the size of the survey sample after it became clear that the response rate would be about 60 to 65 percent (see Appendix B). Furthermore, we released more sample in states with lower response rates than higher response rates, and took into account differential response rates for participants and nonparticipants.

3. ***The sample universe was small in some states.*** In some states, the sample universe was smaller than the size of the worker sample targeted for baseline surveys. In these cases, we selected the entire state universe for the baseline survey sample.

4. ***California specified a maximum sample size.*** Thus, the selected survey sample in California was smaller than the targeted sample size.

5. ***The sampling took into account the likelihood that some nonparticipants would be reclassified as participants.*** To account for these "switchers," we sampled more nonparticipants and slightly fewer participants than equation (2) would suggest, in anticipation that some nonparticipants would be reclassified as participants using baseline survey and more recent TRA benefit receipt data. As discussed, based on the baseline data, the median state switching rate was about 25 percent.

Table A.5 displays the resulting baseline survey sample size, by state and initial TAA participation status. The total survey sample contains 4,381 workers (2,875 participants and 1,506 nonparticipants). The state samples ranged from 110 to 365, with a median state sample size of 141.

TABLE A.5

WORKER COUNTS IN THE SURVEY SAMPLE, BY STATE AND INITIAL TAA PARTICIPATION STATUS

Study State	TAA Participants[a]	TAA Nonparticipants[a]	Total
North Carolina	236	121	357
Pennsylvania	149	73	222
Georgia	124	61	185
Wisconsin	88	44	132
Ohio	106	64	170
California	241	124	365
Tennessee	109	57	166
Illinois	117	58	175
Arkansas	88	48	136
Alabama	92	45	137
Virginia	92	46	138
Michigan	135	67	202
New York	92	46	138
Indiana	97	49	146
South Carolina	125	64	189
Texas	126	57	183
Missouri	93	45	138
Kentucky	92	44	136
Colorado	93	47	140
New Jersey	94	52	146
Maryland	85	47	132
Florida	90	45	135
New Hampshire	79	63	142
Washington	81	53	134
Rhode Island	83	44	127
Minnesota	68	42	110
Total	**2,875**	**1,506**	**4,381**

[a]Participation status was defined using TRA benefit receipt information in the UI Claims files.

METHODOLOGICAL APPENDIX B

THE BASELINE SURVEY

This three-part appendix discusses the design and implementation of the baseline survey as it pertains to the TAA participants and nonparticipants in the survey sample (hereafter referred to as the "treatment group"). First, we discuss the baseline survey design. Second, we discuss survey response rates, and finally, we discuss results from the analysis assessing potential nonresponse bias. The forthcoming report on study implementation discusses these topics in more detail, including a discussion of response rates for the matched comparison group.

1. Design of the Baseline survey

Baseline interviewing took place by telephone between March 2008 and April 2009. Across the 26 study states, 4,381 treatment group members were released for interviews. This sample includes 2,875 TAA participants and 1,506 nonparticipants (using the initial TAA participation designations). An additional 8,875 matched comparison group members were also released for baseline surveys, but these workers are not discussed here.

As discussed in Appendix A, states provided the data necessary for selecting the study samples at different times throughout 2008. Because of the uncertainty as to which states would participate in the study and the dates when they would provide the data, we released workers for baseline surveys in several batches between March 2008 and January 2009 (see Table B.1). The final batch in January 2009 was a supplemental sample across all states that was released to offset the lower-than-expected survey response rates, so that we could complete our targeted number of interviews by the time the survey ended in April 2009. These completion targets were 1,770 for participants and 885 for nonparticipants.

Because sample members were released for surveys on a rolling basis, the length of the follow-up period for locating and interviewing differed somewhat across the sample (Table B.1). This exposure period ranged from three to thirteen months; the exposure distributions are very similar for participants and nonparticipants (Table B.1). As discussed below, survey

TABLE B.1

DISTRIBUTION OF BASELINE SURVEY RELEASE DATES, BY TAA PARTICIPATION STATUS

Release Month and Year (Number of States)	TAA Participants[a] (Percentages)	TAA Nonparticipants[a] (Percentages)	Total (Percentages)
March 2008 (3)	9.0	9.1	9.0
April 2008 (3)	9.3	9.0	9.2
May 2008 (1)	3.1	3.0	3.0
June 2008 (8)	26.3	25.2	25.9
August 2008 (1)	3.7	3.6	3.7
September 2008 (4)	13.6	13.1	13.4
October 2008 (2)	4.9	4.9	4.9
November 2008 (2)	5.5	5.4	5.5
December 2008 (3)	17.8	18.9	18.2
January 2009: Supplemental Sample (26)	6.9	7.8	7.2
Sample Size	**2,875**	**1,506**	**4,381**

[a]Participation status was defined using TRA benefit receipt information in the UI claims files.

response rates increased slowly after three months of exposure. Thus, it is unlikely that the different exposure times across the sample had a large effect on the overall survey response rate.

The UI claims data and certified worker lists provided the contact information for the survey. The UI data contain a telephone number and address for each record. Because sample members sometimes had multiple records in the UI data, we used information on each unique telephone number and address that was available in the data. The certified worker lists also contain an address and sometimes a telephone number for each treatment group member.

The available contact information was somewhat old for some sample members. The median time between the UI claim and interview release date was about 27 months, and the time span was more than three years for about 15 percent of the sample. Thus, UI data on Social Security Numbers (SSNs), names, and dates of birth (which were available for nearly all sample

members) were critical for searching national databases (such as Lexis-Nexis) to help locate sample members who could not be initially reached using the contact information in the UI data and the certified worker lists.

The Office of Management and Budget (OMB) approved the use of incentive fees to all treatment and matched comparison group sample members for completing the survey. However, the structure of the incentive payments changed mid-way during the survey period to help boost response rates. Between March 2008 and mid-September 2008, sample members were offered a $25 incentive for completing the survey.[2] However, for the remainder of the survey period, the incentive increased to $50 for nonparticipants and comparison group members to help increase their response rates; the incentive remained at $25 for participants, whose initial response rates were higher than for the other workers.[3]

The survey questionnaire included a battery of questions about workers' experiences with the TAA program, their labor market and training experiences, and other key study outcomes. The survey coverage period started with the UI claim date associated with the trade-related job separation. The key categories of survey data items were as follows:

[2] More specifically, an experiment was conducted to test the impact of variations in the timing of the incentive payment on response rates. About 60 percent of workers were randomly assigned to receive $25 for interview completion, 20 percent to receive a $2 pre-payment and a $25 interview completion post-payment, and the other 20 percent to receive a $5 pre-payment and a $20 interview completion post-payment (see Gemmill et al. 2009).

[3] These revised incentive payments were approved by OMB based on results from another incentive experiment that was conducted between September and December 2008 with the following features: (1) for participants, 50 percent continued to be eligible for a $25 incentive, and the other 50 percent became eligible for a $50 payment; and (2) for the other groups, 20 percent continued to be offered $25, 40 percent were offered $50, and the final 40 percent were offered $75 (see Schochet et al. 2008 and Gemmill et al. 2009).

- ***Information on the job that led to the UI claim,*** including occupation, industry, union membership, company size, start and end dates, hours worked per week, earnings, available fringe benefits, main reason stopped working, expected recall status, actual recall status, whether looked for work after the job ended, and the number of jobs and total earnings during the prior three years.

- ***Notification of TAA eligibility***, including ways in which found out about TAA, whether an eligibility notification letter was sent by the state and when, whether attended a TAA orientation, and where that meeting took place.

- ***Knowledge of TAA services,*** including knowledge of TRA benefits and TRA eligibility rules (such as having to enter a training program or receive a waiver), knowledge of TAA-funded training and subsidies for travel and relocation, knowledge of ATAA (for those 50 and older), and knowledge of HCTC.

- ***Application for TAA services,*** including whether completed a program application form (and the main reason for applying/not applying), whether applied for ATAA benefits (and, if not, the main reason for not applying), and whether applied for HCTC (and, if not, the main reason for not applying).

- ***The receipt of TRA, ATAA, and HCTC services,*** including TRA benefit receipt information, whether received the ATAA wage supplement and the amount received, and whether received a HCTC tax credit and the amount received.

- ***The receipt of reemployment services,*** including whether received job search assistance, referrals to jobs, resume writing assistance, information on how to change careers, occupational assessment tests, labor market information about what jobs were in demand, information on education or job training programs, and payments for travel, living, and moving expenses; the place where the majority of reemployment services were received; and the helpfulness of the services that were received in finding a job or a training program.

- ***The receipt of education and training services,*** including, for each program: start and end dates, hours per week attended program, type of program, place where the training was received, program cost, sources of funding (including TAA), program completion status, whether received a program credential, and the main reason left the program.

- ***Information on jobs held since the UI claim date,*** including, for each job: occupation, industry, start and end dates, how the job was found, union membership, hours worked per week, earnings, available fringe benefits, reasons stopped working, and the main activity after leaving the job.

- ***Other sources of income,*** including the receipt of public assistance (such as cash assistance and food stamps), pension benefits, and total income from all sources in the prior year.

- ***Household structure,*** including marital status, housing type, household size, and number of children.

- ***Health status and health insurance,*** including whether has a health condition that limits the amount of work that can be performed, the type of health problem, the number of months covered by health insurance since the UI claim date, and the main type of health insurance held by the worker.
- ***Demographic information,*** including education level, race and ethnicity, main language spoken at home, and English language ability.

The companion baseline report focuses on the data items pertaining to the application, knowledge, and receipt of TAA services, the receipt of reemployment services, and the receipt of education and training services. A future report presenting program impact estimates will focus on the full gamut of study outcomes (and, in particular, on the employment-related outcomes).

2. Response Rates to the Baseline Survey

The (unweighted) response rate to the baseline survey for the treatment group was 65.3 percent—68.7 percent for TAA participants and 58.8 percent for TAA nonparticipants (see Table B.2). Overall, interviews were completed with 2,860 of 4,381 treatment group members who were released for interviews. We completed interviews with 1,974 of 2,875 released participants and 886 of 1,506 released nonparticipants. Thus, we achieved our targeted number of completed interviews (1,770 for participants and 885 for nonparticipants). The survey took approximately 38 minutes to complete.

Response rates differed somewhat across key population subgroups (see Table B.2). These subgroups were defined using UI data and local labor market variables that are available for *both* survey respondents and nonrespondents (see Appendix A). Response rates were higher for females than males and increased with age. Response rates were also higher for whites than for other race/ethnicity groups. In addition, response rates were noticeably higher in areas where unemployment and poverty rates were high and in nonmetropolitan areas. In general, the survey response patterns are similar for participants and nonparticipants. A more formal nonresponse analysis is discussed in the next section.

TABLE B.2

RESPONSE RATES TO THE BASELINE SURVEY, BY TAA PARTICIPATION STATUS AND KEY SUBGROUP

Subgroup	Response Rate (Percentages)		
	TAA Participants[a]	TAA Nonparticipants[a]	Combined Sample
Full Sample	68.7	58.5	65.3
Demographic Characteristics			
Age at UI Claim Date			
16 to 40	64.9	54.3	61.1
41 to 50	68.2	56.2	64.1
51 to 60	71.3	66.0	69.6
Older than 60	75.0	63.8	71.4
Gender			
Male	66.5	56.4	62.6
Female	70.7	62.4	68.3
Race/Ethnicity			
White	72.0	65.2	69.6
Black	71.1	56.2	66.5
Hispanic	57.6	45.4	52.8
Other	60.1	45.5	55.2
Benefit Year Start Date			
Before 12/11/05	66.9	52.3	62.3
12/11/05 to 5/28/06	68.3	58.4	65.1
5/28/06 to 10/29/06	70.1	56.4	65.3
Later than 10/29/06	69.4	68.2	68.9
UI Maximum Benefit Amount			
Less than $ 4,524	62.4	54.0	58.4
$4, 524 to $6,048	69.9	56.5	65.8
$6,048 to $7,878	68.3	56.3	64.2
$7,878 to $9,412	68.0	64.0	66.7
$9,412 to $11,700	70.0	61.2	67.2
$11,700 or more	72.3	57.5	66.9
Base Wage			
Less than $ 14,625	68.0	54.3	61.7
$14, 625 to $20,921	68.4	57.3	65.0
$20,921 to $29,520	66.9	54.5	63.0
$29,520 to $42,437	71.4	61.7	68.3
$42,437 to $57,394	66.5	64.6	65.8
$57,395 or more	71.1	60.3	66.1

Table B.2 *(continued)*

Subgroup	Response Rate (Percentages)		
	TAA Participants[a]	TAA Nonparticipants[a]	Combined Sample
Local Labor Market Characteristics			
USDOL Region			
1	66.1	53.2	61.3
2	70.2	59.0	66.5
3	70.2	62.2	67.5
4	63.5	52.0	59.7
5	71.9	66.1	69.9
6	63.7	47.5	57.9
Annual Unemployment Rate (Percents)			
Less than 3.7	58.6	48.2	54.3
3.7 to 4.4	68.2	56.0	64.2
4.4 to 5.1	62.9	57.3	60.9
5.1 to 6.0	73.0	61.2	69.0
6.0 to 7.3	73.7	63.6	70.4
7.3 or higher	75.4	70.8	74.0
2004 Poverty Rate (Percents)			
Less than 7.8	66.6	56.1	62.5
7.8 to 9.8	69.6	57.2	65.3
9.8 to 12.8	64.7	56.3	62.0
12.8 to 15.4	69.6	59.1	65.6
15.4 to 18.0	72.1	62.4	69.3
18.0 or higher	71.9	66.2	70.1
Average Earnings per Job in 2005			
Less than $28,058	79.3	70.3	76.1
$28,058 to $31,760	70.8	64.4	68.8
$31,760 to $38,026	72.9	63.6	69.8
$38,026 to $44,925	67.1	54.0	62.5
$44,925 to $55,716	61.9	55.6	59.6
$55,716 or higher	56.8	44.4	52.6
Percentage of Workers in Manufacturing			
Less than 5.3	62.0	54.1	59.1
5.3 to 7.9	68.9	52.8	63.4
7.9 to 11.2	68.2	61.5	65.9
11.2 to 15.8	68.4	58.6	65.0
15.8 to 21.8	71.3	61.8	68.1
21.8 or higher	73.4	62.7	69.6
Percentage Population Growth Between 2000 and 2005			
Less than -1.9	70.3	61.2	67.0
-1.9 to 0.2	69.7	63.5	67.9
0.2 to 2.8	72.3	62.3	69.2
2.8 to 5.9	66.2	56.4	62.5
5.9 to 12.3	64.2	60.3	62.8
12.3 or higher	67.6	47.2	60.3

Table B.2 *(continued)*

Subgroup	Response Rate (Percentages)		
	TAA Participants[a]	TAA Nonparticipants[a]	Combined Sample
Economic Research Service Urban-Rural Continuum Rating			
Metropolitan area with at least 1 million persons	62.7	52.7	59.3
Metropolitan areas with fewer than 1 million persons	69.5	58.1	65.4
Small area adjacent to a metropolitan area	75.3	64.6	72.1
Small area not adjacent to a metropolitan area	71.9	70.5	71.4
Sample Size	**2,875**	**1,506**	**4,381**

Source: Baseline survey data, UI claims data, and the local area characteristics described in Appendix A.

[a]Participation status was defined using TRA benefit receipt information in the UI claims files.

About 61 percent of all respondents completed interviews within one month after being released for interviewing, and about 85 percent of respondents completed interviews within three months (Table B.3). The distributions of the number of months until completion are similar for participants and nonparticipants. Furthermore, using only sample members who were "exposed" for interviews for at least 8 months, we find that nearly 80 percent of respondents completed interviews within three months (Table B.3). This suggests that response rates did not increase substantially after the first three months of exposure.

TABLE B.3

DISTRIBUTION OF THE NUMBER OF MONTHS BETWEEN THE BASELINE SURVEY RELEASE DATE AND COMPLETION OF THE SURVEY, BY TAA PARTICIPATION STATUS
(Percentages)

	All Treatment Group Members			Those Exposed for Interviewing for at Least 8 Months
Number of Months	TAA Participants[a]	TAA Nonparticipants[a]	Combined Sample	Combined Sample
Less than 1	62.6	58.0	61.2	51.9
1 to 2	17.0	16.8	17.0	18.4
2 to 3	6.5	7.9	7.0	7.5
3 to 6	7.9	10.8	8.8	11.3
6 to 9	3.2	4.6	3.7	6.3
9 to 13	2.7	1.8	2.5	4.7
(Average Months)	1.5	1.7	1.6	2.7
Sample Size	**1,974**	**886**	**2,860**	**2,860**

Source: Baseline survey data.

[a]Participation status was defined using TRA benefit receipt information in the UI claims files.

3. Nonresponse Analysis

Our basic statistical approach for assessing the effects of nonresponse was to compare the characteristics of respondents to nonrespondents using the UI and local labor market data discussed in Appendix A. Table B.4 presents these comparisons in a different way than Table B.2. Table B.4 shows the percentages of respondents and nonrespondents with a particular characteristic (for example, the percentages who are female), whereas Table B.2 displays survey response rates for particular subgroups (for example, for males and females). Another difference between the two tables is that the figures in Table B.4 were computed using sample weights, whereas the Table B.2 figures are unweighted.

We used standard statistical tests to assess the similarity of respondents and nonrespondents, and the statistical significance of these tests is denoted in Table B.4 by asterisks. Using the estimation methods discussed in Appendix C, we used univariate t-tests to compare variable means for binary and continuous variables and chi-square tests to compare variable distributions for categorical variables. In addition, we conducted a more formal multivariate analysis to test the hypothesis that key variable means and distributions are *jointly* similar. For this analysis, we estimated logit regression models where the probability a worker was a respondent versus a nonrespondent was regressed on a set of worker characteristics. Chi-square (log-likelihood) tests were used to assess whether the explanatory variables in the logit models were jointly statistically significant.

There are some differences in the characteristics of baseline survey respondents and nonrespondents that parallel the subgroup differences in response rates that were discussed above (Table B.4). For example, females, whites, and older workers were significantly more likely than their counterparts to complete an interview. In addition, response rates were significantly higher in areas with higher unemployment rates and lower average earnings than in other areas. In

TABLE B.4

COMPARISON OF THE CHARACTERISTICS OF RESPONDENTS AND NONRESPONDENTS TO THE BASELINE SURVEY, BY TAA PARTICIPATION STATUS
(Percentages)

	TAA Participants[a]		TAA Nonparticipants[a]	
Characteristic	Respondents	Nonrespondents	Respondents	Nonrespondents
Demographic Characteristics				
Age at UI Claim Date				
16 to 40	30.4	35.1	31.0	36.9*
41 to 50	31.4	31.8	30.0	34.2
51 to 60	28.5	25.6	30.5	22.9
Older than 60	9.7	7.6	8.5	6.0
(Average age)	46.5	44.7*	46.1	43.6*
Female	54.2	48.1*	42.3	37.8
Race/Ethnicity				
White	64.1	59.9*	68.7	56.2*
Black	22.2	18.6	15.0	16.6
Hispanic	5.7	9.8	8.1	12.2
Other	8.0	11.7	8.2	14.9
Benefit Year Start Date				
Before 12/11/05	20.7	22.7	19.9	23.4*
12/11/05 to 5/28/06	30.8	30.0	28.8	29.6
5/28/06 to 10/29/06	30.5	28.3	26.3	30.0
Later than 10/29/06	18.0	19.1	25.0	17.0
Days Between UI Benefit Year Start Date and the UI First Payment Date				
0 to 5	27.8	32.2	26.3	25.2
6 to 15	30.6	31.4	28.2	25.9
15 to 20	17.6	18.1	19.4	22.4
21 or longer	24.1	18.3	26.1	26.5
Number of UI Records in State Data				
1	49.9	54.1*	71.8	72.9
2	30.1	32.1	22.1	22.6
3 or more	20.0	13.8	6.2	4.5
UI Maximum Benefit Amount				
Less than $ 4,524	7.8	9.8	11.1	12.0
$4, 524 to $6,048	20.8	18.8	13.3	14.4
$6,048 to $7,878	24.1	22.9	21.2	24.5
$7,878 to $9,412	22.7	25.2	33.3	27.7
$9,412 to $11,700	17.6	16.5	15.5	14.9
$11,700 or more	7.0	6.8	5.7	6.6
(Average benefit amount)	7,816	7,783	7,835	7,754

Table B.4 *(continued)*

Characteristic	TAA Participants[a]		TAA Nonparticipants[a]	
	Respondents	Nonrespondents	Respondents	Nonrespondents
Base Wage				
Less than $ 14,625	8.0	8.1	11.9	12.2
$14, 625 to $20,921	17.2	17.1	12.5	13.6
$20,921 to $29,520	27.9	28.3	19.8	24.5
$29,520 to $42,437	26.3	23.7	24.1	21.5
$42,437 to $57,394	13.2	15.9	16.3	15.7
$57,395 or more	7.4	7.0	15.3	12.5
(Average Wage)	32,222	32,046	36,493	34,760
Local Labor Market Characteristics				
USDOL Region				
1	9.4	10.9	6.0	7.2*
2	14.6	14.0	14.1	14.7
3	43.7	40.6	31.3	27.6
4	9.6	9.4	9.4	11.7
5	18.2	19.7	31.9	26.3
6	4.5	5.5	7.4	12.5
Unemployment Rate (Percents)				
Less than 3.7	6.8	10.7*	8.6	12.9*
3.7 to 4.4	16.7	16.3	13.9	16.8
4.4 to 5.1	23.4	28.4	26.8	29.3
5.1 to 6.0	27.7	23.8	27.3	23.2
6.0 to 7.3	16.5	13.8	14.8	12.2
7.3 or higher	8.9	7.1	8.6	5.6
(Average unemployment rate)	5.4	5.2*	5.4	5.1*
2004 Poverty Rate (Percents)				
Less than 7.8	6.8	7.8	9.4	11.3
7.8 to 9.8	11.6	12.2	15.7	17.4
9.8 to 12.8	24.3	26.3	25.4	27.7
12.8 to 15.4	26.6	28.0	25.4	25.4
15.4 to 18.0	19.8	16.9	12.4	11.5
18.0 or higher	11.0	8.9	11.8	6.8
(Average poverty rate)	13.6	13.3	13.0	12.4*
Average Earnings per Job in 2005				
Less than $28,058	12.7	7.5*	12.5	8.9*
$28,058 to $31,760	20.4	19.9	16.7	13.1
$31,760 to $38,026	29.6	25.7	26.7	22.7
$38,026 to $44,925	22.4	26.5	25.0	31.4
$44,925 to $55,716	9.6	10.9	13.6	13.7
$55,716 or higher	5.3	9.5	5.5	10.2
(Average earnings per job)	36,810	39,073*	38,017	40,853*

Table B.4 *(continued)*

Characteristic	TAA Participants[a]		TAA Nonparticipants[a]	
	Respondents	Nonrespondents	Respondents	Nonrespondents
Percentage of Workers in Manufacturing				
Less than 5.3	8.0	9.1	9.2	10.9
5.3 to 7.9	13.0	12.3	12.7	13.6
7.9 to 11.2	19.3	20.3	20.2	20.3
11.2 to 15.8	24.9	26.3	26.0	25.5
15.8 to 21.8	20.2	19.9	18.9	18.4
21.8 or higher	14.7	12.1	13.0	11.4
(Average percentage)	14.0	13.7	13.6	13.1
Percentage Population Growth Between 2000 and 2005				
Less than -1.9	10.2	9.4	12.5	12.1
-1.9 to 0.2	17.6	19.9	15.5	12.8
0.2 to 2.8	29.1	25.1	24.0	21.3
2.8 to 5.9	20.2	22.2	25.5	26.6
5.9 to 12.3	13.5	14.4	13.8	14.4
12.3 or higher	9.4	8.9	8.8	12.9
(Average growth)	3.7	3.7	3.8	4.5
ERS Urban-Rural Continuum Rating				
Metropolitan area with at least 1 million persons	28.7	34.2*	29.7	38.2*
Metropolitan areas with fewer than 1 million persons	31.7	32.2	35.6	34.8
Small area adjacent to a metropolitan area	32.5	26.3	23.9	20.7
Small area not adjacent to a metropolitan area	7.2	7.3	10.9	6.3
Sample Size	**1,974**	**901**	**886**	**620**

Source: Baseline survey data, UI claims data, and the local area characteristics described in Appendix A.

[a]Participation status was defined using TRA benefit receipt information in the UI claims files. All figures are calculated using sample weights.

*Difference between respondents and nonrespondents is statistically significant at the .05 level based on a chi-square test (for categorical variables) or t-tests (for binary or continuous variables).

addition, response rates were significantly higher in rural areas than in larger metropolitan areas. Furthermore, the explanatory variables in the logit models are jointly statistically significant at the 1 percent level for both participants and nonparticipants. Importantly, however, there are no statistically significant differences between the base wages of respondents and nonrespondents.

Because of these differences between the characteristics of respondents and nonrespondents, we adjusted the baseline weights to help reduce the potential bias in the baseline estimates due to survey nonresponse (see Appendix C). The weights were adjusted so that the weighted baseline characteristics of survey respondents were similar, on average, to those of the full population of respondents and nonrespondents. These adjusted weights were used to calculate all statistics presented in the companion baseline report.

It is important to recognize that there may be unmeasured differences between respondents and nonrespondents for which we cannot control. Consequently, our procedure cannot account for the full effects of survey nonresponse. However, because the UI data and local area measures include variables that are likely to be correlated with key study outcomes, we believe that our procedure can account for some important differences between respondents and nonrespondents.

METHODOLOGICAL APPENDIX C

CONSTRUCTION OF SAMPLE WEIGHTS AND STANDARD ERRORS FOR BASELINE ANALYSES

A. INTRODUCTION

This methodological appendix discusses the construction of weights and standard errors for the estimates presented in the companion baseline report. The weights and standard errors both adjust for the sample and survey designs, so that study estimates can be generalized to the certified-worker sample universe.

B. THE CONSTRUCTION OF WEIGHTS

Weights for the baseline analysis were computed for each sample member who completed a baseline survey. These weights were obtained by first calculating the following selection probability for each survey respondent:

$$(1)\quad p_{igs} = q_s * r_{igs} * c_{igs},$$

where p_{igs} is probability that worker *i* in participant group g and state *s* completed a baseline survey; q_s is the probability that state *s* was selected for the study; r_{igs} is the probability that a worker was selected for baseline interviewing among those in the sample universe in state *s*; and c_{igs} is the probability that a worker completed the survey among those released for interviewing. These probabilities were computed separately for participants and nonparticipants (as indicated by the g subscript). The weight for a worker, w_{igs}, was then computed to be proportional to the inverse of the worker's selection probability.

Next, we discuss in turn how we computed each probability in the right-hand-side in (1), and then discuss the construction of the weights and their properties.

1. Computing q_s

The probability that a state was selected for the study was computed using the probabilities displayed in Column 5 of Table A.1 in Appendix A above. These probabilities assume a 26-state design, and are 1 for the 17 certainty states.

As discussed in Appendix A, we randomly selected 25 primary states for the study, and all 25 states ultimately agreed to participate in the study. However, due to the initial reluctance of some states to participate, we contacted several replacement states to increase the chances that we would achieve our target state sample sizes. This process yielded 1 replacement state that agreed to participate in the study, and USDOL decided to include this state in the evaluation. Thus, the final sample includes 26 states.

As shown in Table A.1, the sampling probabilities are very similar using a 25- or 26-state design, and in particular, the two designs yield the same certainty states. Thus, for simplicity, the weights are constructed "assuming" the 26-state design. An alternative approach would be to assign the primary states and the replacement state to different strata and to obtain overall estimates by weighting estimates from each stratum. However, calculating standard errors using this approach would be difficult, because the stratum with the replacement state would have only one state.

2. Computing r_{isg}

The probability that a worker in a particular state was selected for the baseline survey sample was computed by dividing the number of workers released for interviewing in that state by the number of workers in the sample universe for that state. The worker counts in Tables A.4 and A.5 in Appendix A were used for these calculations. However, these counts were adjusted for those who were initially defined as TAA nonparticipants but who were subsequently redefined as TAA participants after they reported in the baseline survey that they had received TAA services. As discussed in Appendix A, the median state "switching rate" was 25 percent, but ranged from 0 to 76 percent across the states. Thus, in each state, we used these switching

rates to update the counts in Tables A.4 and A.5 by increasing participant counts and decreasing nonparticipant counts by the same amount.

3. Computing c_{igs}

Sample members who did not complete a baseline survey may differ from more cooperative sample members who completed the survey in ways that are potentially related to worker outcomes. If not corrected, the effects of interview nonresponse could lead to estimates that might not be generalizable to the study population of eligible TAA workers.

To correct for potential nonresponse bias in the estimates presented in the companion baseline report, we adjusted the sample weights so that the weighted observable baseline characteristics of respondents are similar to the baseline characteristics of the full sample of respondents and nonrespondents. These adjustments were performed using the following three steps:

1. ***We estimated a logit model predicting survey response***. A binary variable indicating whether or not a worker was a respondent to the baseline survey was regressed on state indicators, baseline demographic variables constructed using UI claims data, and local labor market area characteristics (see Table B.4 in Appendix B for a list of the model covariates). A separate logit model was estimated for TAA participants and nonparticipants.
2. ***We calculated a propensity score for each worker in the full sample.*** This score is the predicted probability that a worker was a respondent, and was constructed using the parameter estimates from the logit regression model and the worker's covariate values. Workers with large propensity scores were likely to be respondents, whereas workers with small propensity scores were likely to be nonrespondents.
3. ***We constructed response probabilities (the c_{igs} probabilities) using the estimated propensity scores.*** Workers were ranked by the size of their propensity scores, and divided into five groups of equal size. The response probability for a worker is the mean propensity score of the group to which the worker was assigned.

The selection probabilities, p_{igs}, were then calculated by multiplying estimates of q_s, r_{igs}, and c_{igs}. In addition, for the reasons discussed below, we also computed another set

of selection probabilities, p^*_{igs}, using c^*_{igs} probabilities that were based on logit models that included state indicators only (but no demographic or local labor market area measures).

It is important to note that the propensity score procedure adjusts only for *observable* differences between survey respondents and nonrespondents. The procedure does not adjust for potential unobservable differences between the two groups. Thus, our procedure only partially adjusts for potential nonresponse bias.

4. Constructing w_{isg}

The analysis weights were computed in three stages. First, we calculated initial weights using the relation $w_{igs} = 1/p_{igs}$. Second, we calculated "scaling" weights using the relation $w^*_{igs} = 1/p^*_{igs}$. Finally, we scaled the initial weights so that their sum would equal the sum of the w^*_{igs} weights within each state. We scaled the weights in this way so that state survey response rates would play a major role in the nonresponse adjustments. Under this scheme, corrections for differential response rates across demographic and local labor market area groups were performed within states, not between states.

The resulting weights sum to 27,492 workers for the 2,228 TAA participants in the analysis sample and to 27,155 workers for the 632 nonparticipants in the analysis sample (Table C.1).[4] The median weight is 10 for participants and 33 for nonparticipants. In addition, the interquartile range for the weights is about 11 for the participants and 37 for the nonparticipants (Table C.1).

[4] These universe counts are slightly different than those discussed in Appendix A (26,889 participants and 28,033 nonparticipants) because the Appendix A figures were calculated under the assumption that states were sampled with replacement.

TABLE C.1

DISTRIBUTION OF WEIGHTS, BY TAA PARTICIPATION STATUS

Statistic for Weights	TAA Participants[a]	TAA Nonparticipants[a]
Sum	27,492	27,155
Maximum	42.2	238.7
75th Quantile	17.7	57.7
Median	10.4	33.1
25th Quantile	6.5	20.4
Minimum	3.5	5.5
Sample Size	**2,228**	**632**

[a]Participation status was initially defined using TRA benefit receipt information in the UI claims files, and was then updated using baseline survey information on the receipt of TAA services.

The design effect due to weighting is about 1.29 for baseline analyses that used the participant sample, and about 1.52 for baseline analyses that used the nonparticipant sample. The design effect is 2.12 for analyses that combined the two samples.

5. Adjusting for the Curtailed Two-Year Post-Certification Coverage Period

As discussed in Appendix A, the petition certification period for the study was between November 1, 2005 and October 31, 2006. Workers covered by a certification include those laid off between one year prior to the petition filing date and two years after the petition certification date. Thus, the sample frame for the study includes TAA-eligible workers who received UI benefits between September 1, 2004 and October 31, 2008.

The UI claims data that we received from states typically cover the 2004 to 2007 period, although the coverage period ended in 2006 or mid-2008 for a few states. Thus, the UI data cover all workers who were laid off before the petition filing date and most workers during the two-year period after the certification date. For example, the UI data cover 17 months of the 24-month post-certification period for the average petition. Furthermore, the UI data cover at least

half of the 24-month post-certification period for three-quarters of the petitions (see Table A.3 in Appendix A).

UI coverage rates, however, differ somewhat across states due to differences in the dates that the states extracted the data. Furthermore, the sample under-represents those laid off near the end of their firm's certification window. Thus, we constructed weights to adjust for these sources of underrepresentation in the sample.

In order to construct these weights, we first examined the distribution of the number of months between each worker's UI claim date and their firm's petition certification date. Column 2 of Table C.2 displays an estimated *population* distribution using only those workers whose associated 24-month post-certification window was fully covered by the UI data. Column 3 of Table C.2 displays the *sample* distribution using all workers in the sample.

TABLE C.2

DISTRIBUTION OF THE NUMBER OF MONTHS BETWEEN THE UI CLAIM AND PETITION CERTIFICATION DATE

Number of Months Between the UI Claims and the Petition Certification Date	Sample Whose Certification Window Was Fully Covered by the UI Data	Full Sample
Less than -6	3.9	11.9
-6 to<0	34.5	32.3
0 to 6	34.9	39.1
6 to 12	17.4	12.6
12 to 18	6.1	3.4
18 to 24	3.2	0.7
Sample Size	**507**	**4,307**

[a]Participation status was initially defined using TRA benefit receipt information in the UI claims files, and was then updated using baseline survey information on the receipt of TAA services.

Several key points emerge from Table C.2. First, about 70 percent of UI claims in the population were filed during the six months before and the six months after the certification date. Second, the main source of underrepresentation in the sample are workers who started their UI spells more than 12 months after the petition certification date; these workers constitute about 10 percent of workers in the population, but only about 4 percent of workers in the sample.

Accordingly, we constructed weights to adjust for this underrepresentation by multiplying the baseline weights by (10/4) for workers in the sample who started their UI spells more than 12 months after the petition certification date. This weighting scheme assumes that workers in the sample who were laid off late in the certification window are representative of all such workers in the population.

The estimates presented in the companion baseline report are very similar using these adjusted weights and the unadjusted baseline weights. Thus, for simplicity, the baseline report presents results using the unadjusted weights.

B. THE CALCULATION OF STANDARD ERRORS

The companion baseline report presents estimates of variable means and distributions for the full sample of TAA participants and nonparticipants, and for various population subgroups. The report also presents estimated parameters from several regressions to examine the association between model covariates and key TAA-related outcomes. These estimates were all obtained using SUDAAN and the sample weights discussed above.

This section provides details on the design-based mathematical formulas that were used to obtain these estimates and their standard errors. The standard errors of all estimates presented in the companion report account for design effects due to weighting and state-level clustering.

1. Estimators for Variable Means and Standard Errors

In this section, we discuss the estimation of variable means and their standard errors using the participant sample. The same approach was used to produce statistics for the nonparticipant sample and the combined sample.

As discussed, the design for the TAA evaluation design is a two-stage stratified design, where n_h states (PSUs) were selected within region (strata) h with probabilities proportional to estimated size, and m_{hs} participants were then selected from region-h state-s. Let the weights for worker i be denoted by w_{hsi}.

Under this design, the SUDAAN estimate of the mean for a continuous or binary outcome, y, was calculated as follows:

$$\bar{y} = \frac{\sum_{h=1}^{H}\sum_{s=1}^{n_h}\sum_{i=1}^{m_{hs}} w_{hsi} y_{hsi}}{\sum_{h=1}^{H}\sum_{s=1}^{n_h}\sum_{i=1}^{m_{hs}} w_{hsi}},$$

where all terms were defined above.

SUDAAN uses the Taylor linearization method to calculate the variance of $\bar{y}$. To highlight the features of this method, suppose that we are interested in estimating the variance of a population parameter $\beta=F(x_1,x_2,\ldots,x_n)$ for some function $F(.)$, where the vector x_i is the observed data vector for the ith unit in the sample. Suppose next that we perform a Taylor expansion of β around $(\mu_1,\mu_2 \ldots, \mu_n)$ where $\mu_i = E(x_i)$, where the $E(.)$ operator is the expected value of x_i averaging over repeated sampling from the sample universe. This Taylor expansion yields the following expression for the variance of β:

(3) $\text{var}(\beta) \approx \text{var}(\sum_i Z_i)$, where

$$Z_i = \frac{\partial F}{\partial x_i}(\mu_1, \mu_2 \ldots, \mu_n) x_i.$$

Consequently, to estimate the variance of β, the linearized covariates, Z_i, can be used in formulas for calculating variances for population *totals* under clustered designs.

To use this method for $\bar{y}$, we note that the mean outcome defined above is a *ratio* of two sums (denoted by *R*). Using equation (3), the corresponding linearized variables for this ratio estimator can be expressed as follows:

$$(4)\quad Z_{hsi} = \frac{w_{hsi}(y_{hsi} - \hat{R})}{\sum_{h=1}^{H}\sum_{s=1}^{n_h}\sum_{i=1}^{m_{hs}} w_{hsi}}.$$

As discussed next, the way in which the study used these linearized *Z* variables in the variance calculations differed for those in the certainty and noncertainty states.

Certainty States. As discussed above, 17 states were selected with certainty (because these states had state selection probabilities greater than 1). The worker samples in each of these states were treated as a simple random sample from each state. This is because the certainty states were not "sampled," and hence, each certainty state is effectively its own stratum. Consequently, the variance of a mean outcome in the certainty states does not need to account for between-state variability but only within-state variability.

The study estimated the variance of a mean outcome in the certainty states as follows:

$$(5)\quad \text{var}(\bar{y}_{\text{Certainty}}) = \sum_{h}\sum_{s} m_{hs} S_{hs}^2, \text{ where}$$

$$S_{hs}^2 = \sum_{i=1}^{m_{hs}} (Z_{hsi} - \bar{Z}_{hs})^2 / (m_{hs} - 1)$$

$$\bar{Z}_{hs} = \sum_{i=1}^{m_{hs}} Z_{hsi} / m_{hs}.$$

It is important to note that, for simplicity, the formulas are not indexed by "*certainty*," although this index is implied, because these calculations were performed using data on only those workers in the certainty states. This convention is followed for the remainder of this section.

Noncertainty States. The variances of the estimated means in the 9 noncertainty states must account for clustering due to the sampling of states. The study calculated these variances assuming that the sampling of states was performed with replacement (WR) and using the following variance formula:

(6) $\text{var}(\bar{y}_{\text{Noncertainty}}) = \sum_{h} n_h S_h^2$, where

$$S_h^2 = \sum_{s=1}^{n_h} (Z_{hs} - \bar{Z}_h)^2 / (n_h - 1)$$

$$Z_{hs} = \sum_{i=1}^{m_{hs}} Z_{hsi}$$

$$\bar{Z}_h = \sum_{s=1}^{n_h} Z_{hs} / n_h .$$

In practice, four regions had only one noncertainty state, making it difficult to estimate a between-state variance within these regions. Thus, we estimated a variant of equation (6) where all noncertainty states were assigned to the same stratum.

Finally, it is important to note that very similar standard errors were found assuming a without replacement (WOR) design.

Combining the Variance Estimates. The study calculated overall variance estimates by combining the variance estimates from the certainty and noncertainty states as follows:

(7) $\text{var}(\bar{y}) = p_c^2 \, \text{var}(\bar{y}_{\text{Certainty}}) + (1 - p_c)^2 \, \text{var}(\bar{y}_{\text{Noncertainty}})$,

where p_c is the population share in the certainty states (which is about 80 percent).

Similar methods were used to calculate statistics for nonparticipants, for the combined participant-nonparticipant sample, and for population subgroups. Furthermore, to gauge whether differences in the estimated means between participants and nonparticipants (or between subgroup levels) were statistically significant, we conducted *t*-tests by dividing the differences in

the estimated means by the square root of their estimated variances, or by using *F*-tests to test for subgroup interaction terms.

2. Regression Estimators

For several key dependent variables, the companion baseline report presents parameter estimates from variants of the following regression model:

(8) $y = \alpha + Q\delta + \varepsilon,$

where y is the dependent variable, Q are (k-1) explanatory variables, ε is a mean zero disturbance term, and α and δ are model parameters.

The study used the generalized linear model procedures in SUDAAN to estimate the regression parameters and their variances. These methods generalize the Taylor series linearization method discussed above for parameters that are defined as *implicit* functions of linear statistics or estimating equations. These methods can be used to estimate linear models for continuous outcome measures as well as nonlinear logistic models for binary outcomes.

The theoretical assumptions for generalized linear models are as follows:

(9) $E(y_{hsi}) = \mu_{hsi},$

(10) $Var(y_{hsi}) = Var(\mu_{hsi}),$

and g is a link function such that:

(11) $g(\mu_{hsi}) = x'_{hsi}\beta$ and $\mu_{hsi} = g^{-1}(x'_{hsi}\beta).$

Note that the X variables in equation (11) contain both the intercept and Q variables in equation (8), and that the $kx1$ parameter vector β contains both the α and δ parameters.

The estimating equations for the exponential family of distributions (of which linear and logistic regressions are special cases) can be derived by setting to zero the derivatives of the log likelihood function with respect to β. These estimating equations can be expressed as follows:

$$(12)\quad \frac{\partial \log L}{\partial \beta} = S(\beta) = \sum_{h=1}^{H}\sum_{s=1}^{n_h}\sum_{i=1}^{m_{hs}} \frac{\partial \mu_{hsi}}{\partial \beta} w_{hsi} V(\mu_{hsi})^{-1}(y_{hsi} - \mu_{hsi}) = 0,$$

where $S(\beta)$ is the score function.

Estimates of β in equation (13) can be obtained using Newton-Raphson (Taylor Series) methods. The variance of these estimates can be calculated as follows:

$$(13)\quad \text{var}(\hat{\beta}) = (J_0)^{-1} Var[S(\hat{\beta})](J_0^{/})^{-1},$$

where J_0 is a *k*-by-*k* matrix of derivatives of the score function with respect to β, and $Var[S(\beta)]$ is the *design-based* variance of the score function.

An estimate of $Var[S(\beta)]$ can be obtained using the Taylor linearization method discussed in the previous section. This is because the score function is a *sum* of linearized *Z* vectors, where the Z vector for each individual is of the form:

$$(14)\quad Z_{hij} = \frac{\partial \mu_{hij}}{\partial \beta} w_{hij} V(\mu_{hij})^{-1}(y_{hij} - \mu_{hij}).$$

Consequently, similar procedures to those described in the previous section for the estimated means can be used to compute $Var[S(\beta)]$ using the linearized Z vectors. For instance, the variance estimate in the noncertainty states can be computed as follows:

$$(15)\quad Var[S(\hat{\beta})] = \sum_h \frac{n_h}{n_h - 1} \sum_s (Z_{hs} - \bar{Z}_h)(Z_{hs} - \bar{Z}_h)^{/}$$

$$Z_{hs} = \sum_i Z_{hsi}$$

$$\bar{Z}_h = \frac{1}{n_h} \sum_s Z_{hs}.$$

Linear and logistic regression procedures are special cases of the above generalized linear model formulation. For linear regression, the β parameters can be estimated using the following weighted least squares formula:

(16) $\hat{\beta} = (X^{/}WX)^{-1}X^{/}WY,$

where W is a matrix of weights. Design-based variances for these regression coefficients can be estimated using the formulas in equations (12) to (14) where:

(17) $\mu_{hsi} = x_{hsi}^{/}\beta$ and $Var(\mu_{hsi}) = \sigma^2.$

For logistic regression models, the assumptions are:

(18) $\mu_{hsi} = \dfrac{\exp(x_{hsi}^{/}\beta)}{1+\exp(x_{hsi}^{/}\beta)}$ and $Var(\mu) = \mu(1-\mu).$

REFERENCES

Area Resource File (2007). U.S. Department of Health and Human Services, Health Resources and Services Administration.

Inter-University Consortium for Political and Social Research (ISPSR) Data File (2007). University of Michigan, Institute for Social Research.

Bureau of Labor Statistics, United States Department of Labor (2007). Local Area Unemployment Statistics.

Gemmill, E., P. Nemeth, P. Schochet, J. Berk (2009). "How Procedural and Incentive Payment Changes Can Increase Response Rates." Princeton, NJ: Mathematica Policy Research Working Paper.

Schochet, P., J. Berk, P. Nemeth (2008). "Short-Term Results of the New Survey Procedures for the TAA Evaluation." Princeton, NJ: Mathematica Policy Research Study Memo.