

**Evaluation of the Unemployment
Compensation Provisions of the
American Recovery and
Reinvestment Act of 2009**

**OMB Supporting Statement:
Part B**

March 13, 2012

Authors (in alphabetical order):

Heinrich Hock

Brandon Kyler

Annalisa Mastri

Julita Milliner-Waddell

Karen Needels

Patricia Nemeth

Walter Nicholson

Frank Potter

Grace Roemer

Linda Rosenberg

Wayne Vroman*

*The Urban Institute

MATHEMATICA
Policy Research, Inc.

Contract Number:
GS10F0050L/DOLF109631341

Mathematica Reference Number:
06863.450

Submitted to:
U.S. Department of Labor
Office of the Chief Evaluation Officer
200 Constitution Avenue NW
Washington, DC 20210
Project Officer: Jonathan A. Simonetta

Submitted by:
Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005
Project Director: Karen Needels

**Evaluation of the Unemployment
Compensation Provisions of the
American Recovery and
Reinvestment Act of 2009**

**OMB Supporting Statement:
Part B**

March 13, 2012

Authors (in alphabetical order):

Heinrich Hock
Brandon Kyler
Annalisa Matri
Julita Milliner-Waddell
Karen Needels
Patricia Nemeth
Walter Nicholson
Frank Potter
Grace Roemer
Linda Rosenberg
Wayne Vroman*

*The Urban Institute

MATHEMATICA
Policy Research, Inc.

CONTENTS

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS.....	1
1. Respondent Universe and Sampling	2
2. Analysis Methods and Degree of Accuracy	14
3. Methods to Maximize Response Rates and Data Reliability	30
4. Tests of Procedures or Methods	38
5. Individuals Consulted on Statistical Methods	39
REFERENCES	41
APPENDIX A: UI RECIPIENT SURVEY	
APPENDIX B: SURVEY OF UI ADMINISTRATORS	
APPENDIX C: SITE VISIT PROTOCOL	
APPENDIX D: DATA SYSTEMS SURVEY	
APPENDIX E: 60-DAY FEDERAL REGISTER NOTICE	
APPENDIX F: ADVANCE MATERIALS	

TABLES

B.1 Selection Probabilities Based on the UCP and UCP-COBRA Size Measures..... 9

B.2 States' Characteristics by Receipt of UC Modernization Funds Received (as of December 2010)..... 14

B.3 Estimation of Impacts 20

B.4 Design Effects and Effective Sample Sizes for the Two-Stage UI Recipient Sample 27

B.5 Minimum Detectible Subgroup Differences 29

PART B: COLLECTION OF INFORMATION INVOLVING STATISTICAL METHODS

The U.S. Department of Labor (DOL) contracted with Mathematica Policy Research to conduct an evaluation of the unemployment compensation (UC) provisions of the American Recovery and Reinvestment Act (ARRA) of 2009 and related legislation. The major provisions can be grouped into three categories. The first includes extensions of the number of weeks of unemployment benefits available to workers who exhausted their entitlement to state-financed benefits. The Emergency Unemployment Compensation Act of 2008 (EUC08), initially signed in June 2008 but extended several times, contains four tiers of benefits, which collectively can provide up to 53 weeks of additional UC benefits to workers who exhausted their entitlements under regular state UI programs. Legislation also made additional changes to expand the availability of benefits through the Extended Benefits (EB) program, a long-standing program that provides additional weeks of benefits to unemployed workers in states with unemployment rates above certain thresholds. Furthermore, the EB program, which historically had been financed 50-50 by states and the federal government, could be fully financed by the federal government. The second category of UC provisions is intended, through the use of federal incentive funds offered to states, to encourage states to modernize their programs in response to certain changes over time in technology and the labor market. The policies have the intent of expanding UC system coverage to additional workers or providing adding benefits to covered workers. The third set of provisions is intended to help states or unemployed workers better weather the recession. These provisions include (1) the establishment of Federal Additional Compensation (FAC), which added \$25 per week to UC weekly benefit amounts until it expired in December 2010; (2) a reduction in federal taxation of a portion of UC benefits during calendar year 2009; and (3) suspension of interest payments on all state trust fund loans in 2009 and 2010. The net result of these changes and other UC-related provisions of ARRA was that the federal government came to play a much larger role in the UC system than had been the case in previous recessions.

The evaluation of the UC provisions of ARRA and the related legislation is designed to provide insights about five topic areas: (1) states' decisions to adopt certain UC-related reforms encouraged by ARRA, (2) states' implementation experiences with ARRA UC provisions, (3) the characteristics of recipients of different types of unemployment benefits during the time in which ARRA-related UC benefits were available, (4) impacts of UC ARRA provisions on recipients' outcomes, and (5) additional research questions about the influence of the UC provisions of ARRA on macroeconomic issues and state unemployment insurance (UI) trust funds.

This package requests clearance for three data collection efforts conducted as part of the evaluation:

1. ***A UI Recipient Survey.*** This survey will seek information about a nationally representative sample of approximately 2,400 UI recipients in 20 randomly selected UI jurisdictions from among the states and the District of Columbia; topics to be covered include the recipients' employment and financial characteristics prior to their unemployment spells, as well as their experiences during and after benefit collection. The UI recipient survey is presented in Appendix A.
2. ***A Survey of UI Administrators.*** This survey will yield data about the decision-making and implementation experiences of UI administrators in all 50 states and the District of Columbia. The survey of UI administrators is presented in Appendix B.

3. **Site visit Data Collection.** In-person visits to 20 purposively selected states and a data systems survey to be provided to state-level staff prior to the in-person visits will provide qualitative and in-depth information about the states' experiences deciding whether to adopt optional UC-related provisions of ARRA, as well as their experiences with implementation of these and other provisions. A master protocol for the visits and the data systems survey are included in Appendixes C and D, respectively.

1. Respondent Universe and Sampling

The following three subsections discuss the respondent universe and sampling for the UI recipient survey, the survey of UI administrators, and the site visit data collection, respectively.

a. UI Recipient Survey

The individual-level analyses conducted for this study were commissioned by DOL to determine how the experiences of job losers were shaped by the modifications to the UC system enacted by the federal government in response to the recent recession. The study's impact evaluation seeks to measure the effects of certain ARRA-based changes to UC policies (for example, availability of extended benefits for UI exhaustees through the four tiers of the EUC08 program) on labor market, training, and financial outcomes of UI recipients. Key study outcomes include the duration of the initial unemployment spell, earnings on reemployment, and the extent of financial hardships that recipients experienced. (A more detailed list of the UC policy changes considered in the impact analysis is included in Section B.2.) To put the impact estimates in context, descriptive analyses will also provide DOL with an understanding of the socioeconomic and demographic characteristics of unemployed workers served by the UC system during the recent recession. Because most of these characteristics and outcomes are either imperfectly measured or not measured at all in administrative and extant survey data, Mathematica will conduct a survey of UI recipients to gather the unique data needed for this evaluation.

To cost-effectively produce nationally representative estimates, the survey will be administered to a sample of 2,400 UI recipients identified from administrative claims records using a two-stage cluster randomized sampling strategy. In the first stage, a sample of 20 out of the 51 major *UI jurisdictions* (50 states and the District of Columbia) will be randomly selected from which to gather the administrative data to identify recipients (the "sampling frame"). In the second stage, 3,000 recipients from the jurisdictions selected in the first-stage sample will then be randomly selected to be interviewed. An expected individual-level response rate of 80 percent will yield a sample of 2,400 recipients completing surveys.¹ Although these sample sizes were limited by budgetary considerations, they should be sufficient to measure differences between important study subgroups with reasonable precision. For example, as described in Section B.2, power calculations based on this sampling design will allow differences in the gender composition of UI-only recipients and extended-benefits recipients of between 6.6 and 8.0 percentage points to be detected. Moreover, when comparing recipients who experienced a gap between UI exhaustion and the availability of extended benefits to recipients who were able to progress smoothly from UI benefits onto EUC08 benefits, the survey sample is expected to yield a minimum statistically detectable difference in

¹ Section B.3 describes analyses and adjustments that will be made to address the potential for non-response bias at both the individual level and at the state level, as well as methods that will be used to maximize response rates.

unemployment durations of 3 months. Nationally representative sample statistics will be estimated using weights that are derived from the sampling design.

Although the two-stage sampling design will result in less precise estimates than what would be obtained if recipients were interviewed from every UI jurisdiction, it substantially reduces both the resources required for the study and the collective burden to be incurred by UI jurisdictions in providing the administrative files. As a further measure to minimize burden and costs incurred from UI records that will serve as the sampling frame, states will be selected jointly for this study (the “UCP Study”) and Mathematica’s DOL-funded evaluation of the COBRA subsidy available under ARRA (the “COBRA Subsidy Evaluation”).²

The following subsections describe specific elements of the two-stage sampling strategy in greater detail: (1) the target population and study populations for this evaluation; (2) the allocation of the second-stage survey sample across benefit year begin (BYB) dates and study subpopulations; (3) the “composite size measure” that is used to calculate first-stage selection probabilities for UI jurisdictions; (4) the stratification system that, in conjunction with selection probabilities, determines the likely distribution of UI jurisdictions included in the sample; and (5) the sampling weights that will be constructed to account for the sampling design.

1. Target Population and Study Population

The overall target population for the evaluation consists of individuals who were potentially eligible for additional unemployment benefits through the EUC08 legislation. Thus, recipients with BYB dates ranging from May 1, 2006, through late 2011 (given current legislation) could potentially be included in the analysis.

The survey will concentrate on a study population with BYB dates between October 1, 2007, and September 30, 2009. This range of BYB dates includes recipients with diverse experiences with ARRA-related policy and program changes related to the duration of benefits available. For example, most of the variation in the number of weeks of benefits that could be collected without interruption through the EUC08 and EB programs applied to individuals with BYB dates in 2008. Thus, concentrating the survey sample on recipients with BYB dates ranging from the last quarter of 2007 to third quarter of 2009 will result in more precise estimates of the impact of the benefits available under EUC08 and EB, as compared to a broader date range that includes earlier and more recent recipients. It also allows the full UC benefit collection history to be characterized for most survey respondents using administrative data, reducing the need to ask for this information in the survey or to use statistical techniques to account for incomplete information. Finally, post-UC outcomes will be observed for most recipients in the survey, which will increase the capacity of the evaluation to detect impacts.

² As with the UCP study, the COBRA Subsidy Evaluation seeks to implement a two-stage cluster randomized design with 20 UI jurisdictions selected in the first stage. The COBRA study will focus on a study population consisting of UI recipients who lost their jobs between February 17, 2009, and May 31, 2010, drawing a sample of 12,000 individuals to be located for interviewing. A separate OMB/PRA clearance package will be submitted for data collection for the COBRA Subsidy Evaluation.

As described in Section B.2, the descriptive and impact analyses will estimate and compare the average characteristics and outcomes of various subgroups of UI recipients. One key comparison for this study will be between the following subpopulations:

- The UI-only population, consisting of individuals who received a first payment from the regular UI system and did not exhaust their regular UI entitlement; and
- The extended benefits population, consisting of individuals who received a regular UI first payment and subsequently collected benefits through EUC08, through EB, or through both programs.

These subpopulations partition the overall study population described above. The extended benefits population does not distinguish among individuals according to the program from which their benefits were derived. The reason for this is that there is potentially substantial overlap in duration of benefits between recipients of EUC08 Tier 4 benefits and EB recipients in UI jurisdictions that had not triggered onto Tier 4.³ Hence, the experiences of those two groups of recipients might be fairly similar. Furthermore, some jurisdictions transitioned UI exhaustees onto EB during lapses in the EUC08 legislation; in these cases, recipients would not progress in a clean, sequential way through the three types of programs.

2. Allocation of the Second-Stage Survey Sample Across BYB Dates and Study Populations

The sample of recipients will be allocated fairly equally across six-month ranges of BYB dates.⁴ This should allow the effects of EUC08 on recipient outcomes to be detected with more reliability, as compared to a proportional allocation across BYB dates that would tend to arise naturally with unrestricted sampling. Many of the changes to EUC08 legislation affected individuals based on their date of entry into the UI system. For example, workers who continuously collected benefits from a 26-week regular UI entitlement with a BYB date in July 2008 would typically have experienced a three-month gap between when their EUC08 Tier 2 benefits were exhausted and when they became eligible to collect EUC08 Tier 3, whereas workers in a similar situation but with BYB dates in October 2008 would have transitioned smoothly onto Tier 3. Thus, a nearly equal allocation of the sample across BYB date ranges will increase the precision for detecting differences in the effects of the availability of Tier 3 benefits. Greater statistical power for the impact analysis through a disproportionate sample allocation may come at the cost of lower overall descriptive power.⁵ Nonetheless, as shown below, the sample allocation should still yield fairly precise survey-based descriptive statistics.

To achieve an approximately equal number of survey respondents in each BYB date range, selection of UI recipients will be *explicitly stratified* across BYB date range strata. (Within each BYB date stratum, the sample will be allocated across UI jurisdictions to achieve approximately equal

³ Tiers 1, 2, and 3 of EUC08 became available in almost every UI jurisdiction. However, only 33 jurisdictions triggered onto Tier 4 of EUC08.

⁴ Between-group comparisons are generally the most precise when there are equal numbers in the groups being compared.

⁵ A proportional allocation will result in nearly equal weights when generating survey estimates that are representative of the underlying population. Unequal weighting will tend to increase the sampling variance.

sampling weights for sample members in each BYB date range stratum.) Selection will occur independently in each sampling stratum defined by UI jurisdiction and BYB date range stratum. Survey monitoring costs increase with the number of sampling cells and nonresponse analyses become less reliable as the number of sample members in each cell decreases. Thus, it is unworkable to explicitly define very fine-grained BYB date range strata. For example, given the two-year sampling frame and a total of 2,400 surveys to be completed, explicitly stratifying by month would result in 480 sampling cells to monitor (20 UI jurisdictions \times 24 month strata) that would each contain, on average, a target of 5 survey completes. Instead, BYB will be stratified based on six-month intervals. This would result in 80 cells, each containing a target of 30 survey completes—numbers that are more feasible for monitoring and nonresponse analyses.⁶

Within the explicit BYB date range strata, selection of recipients will be *implicitly stratified* according to the BYB month and then by study population within BYB month. Implicit stratification involves first dividing the population list into strata and then sorting records within each stratum by the implicit stratification factors (in this case, by BYB month and study population). The sample is then drawn from this sorted list using a sequential selection procedure (Chromy 1979). Implicit stratification will result in an approximately proportional allocation across BYB months without imposing fixed sample sizes in each stratum, as with explicit stratification, and thus is expected to have less of an effect on monitoring costs.

The decision to implicitly stratify the survey sample by study populations is based on three considerations. First, although an equal allocation maximizes the precision of comparisons *between* groups, a proportional allocation reduces the variation in the sampling weights when computing *pooled* estimates across groups. The latter would increase the precision of pooled analyses of, for example, the relationship between financial well-being (or the duration of benefit receipt itself) and the availability of EUC08 among all UI recipients. Second, the precision loss for between-group comparisons is unlikely to be substantial because preliminary estimates indicate that roughly half of UI recipients moved onto EUC08. Third, an equal allocation across study populations would require explicit stratification, which would double the number of sampling cells and approximately halve the number of cases in each cell. An approximately proportional allocation can be achieved through implicit stratification, which leaves monitoring costs and the reliability of nonresponse analyses unchanged.

3. Composite Size Measure and First-Stage Selection of UI Jurisdictions

The study will select 20 UI jurisdictions in the first stage with probability proportional to a composite size measure defined as a weighted sum of the total population in each explicit second-stage BYB stratum. This composite size measure is calculated as the expected sample size across all study populations in the first stage sampling unit (the UI jurisdictions), as described below. Specifically, states will be selected with probability proportional to a composite size measure that is based on the number of UI recipients who receive first payments in each of four the six-month BYB date ranges described above. This composite size measure also permits sample sizes to be similar across the selected states while minimizing variation in selection probabilities among individuals

⁶ Although it will not be possible to stratify sampling within each BYB date range stratum by additional socioeconomic factors that have been shown to have a significant association with survey response rates, information can be pooled across strata to analyze and adjust for such patterns. More details on the study's nonresponse analyses are provided in Section B.3.

within the same study population. In this design, every recipient within each BYB date range has an equal probability of being included in the final sample, which reduces the losses in precision arising from variation in the sampling weights.

First-Stage and Second-Stage Sample Sizes. The number of states to be included in the study has been determined by two factors. First, although collecting data from all 51 UI jurisdictions would improve precision by avoiding the use of a clustered sampling design, the intensive recruitment efforts and cost-recovery payments required to do so would be prohibitively expensive, given the available resources. Second, because of constraints on the budgetary resources available for individual-level data collection, there is a tradeoff between the gain in precision from increasing the number of states and the loss in precision from a smaller sample size of individual recipients. Based on the past experience of DOL and Mathematica in conducting similar large-scale surveys, sampling 20 jurisdictions is expected to maximize overall precision: the precision gained by including additional jurisdictions is likely to be outweighed by the precision lost due to the smaller sample size. Given the budgetary allocation for individual-level data collection and an expected response rate of 80 percent, selecting 20 UI jurisdictions in the first stage implies that it will be feasible to obtain 2,400 completed surveys based on an initial sample of 3,000 recipients selected from the first-stage jurisdictions' administrative records.

Composite Size Measure. For the UCP study, the composite size measure for each UI jurisdiction j would be set equal to

$$(1) \quad S_j^{UCP} = \sum_{i=1}^4 f_i^{UCP} \times N_{ij}^{UCP},$$

where i indexes each six-month period between October 1, 2007, and September 30, 2009, N_{ij}^{UCP} is the number of UI first payments made in jurisdiction j during period i and f_i^{UCP} is the sampling rate of the national (51-jurisdiction) population of recipients with first payments in period i that will be contacted for interviews. Since approximately 3,000 recipients will be contacted for the study as a whole and that the survey sample will be allocated evenly across six-month intervals:

$$(2) \quad f_i^{UCP} = 750 / N_{i,tot},$$

where $N_{i,tot}$ represents the national number of UI first payments received during period i . The composite size measure, S_j^{UCP} , can be interpreted as the expected number of individuals that would be sampled from jurisdiction j if all jurisdictions were included in the sample with certainty.

To coordinate the selection of UI jurisdictions with the COBRA Subsidy Evaluation, the composite size measure is expanded so that for jurisdiction j it is equal to

$$(3) \quad S_j^{JOINT} = \left[\sum_{i=1}^4 f_i^{UCP} \times N_{ij}^{UCP} \right] + f^{COBRA} \times N_j^{COBRA}.$$

This joint size measure contains all of the components of the UCP composite size measure for jurisdiction j in equation (1) and adds a final term based on the national sampling rate sought for the COBRA study (f^{COBRA}) applied to a count of individuals who received a first UI payment in

jurisdiction j between February 17, 2009, and May 31, 2010 (N_j^{COBRA}). Hence, using the joint size measure puts slightly more weight on individuals who lost their jobs during the trough of the recession and the subsequent recovery period than if the measure were constructed for UCP alone. However, the added COBRA component of the joint size measure is highly correlated with the UCP-alone size measure ($\rho = 0.991$), so adopting the joint measure will have little effect on the jurisdiction-level selection probabilities.⁷

Among the 20 UI jurisdictions to be selected for the study, a few jurisdictions with the largest numbers of UI recipients, as gauged by their composite size measures, will be selected with certainty. These jurisdictions would appear in every random sample that could be drawn and would, on average, be included at least once if the sample were drawn with replacement. The remaining jurisdictions, known as noncertainty jurisdictions, will be selected without replacement using a sequential selection probability proportional to size (PPS) procedure (Chromy 1979) and using the stratification system described in the next subsection.

Allocation of the Second-Stage Survey Sample Across UI Jurisdictions. Conditional on a UI jurisdiction being included in the selected sample, the number of sample members allocated to each BYB stratum in that jurisdiction will vary in proportion to the expected number of individual recipients in that stratum, as compared to the total number of recipients selected in jurisdiction j with first-payment dates in the four BYB strata. More specifically, the number of individuals with BYB dates in period i selected to be interviewed in jurisdiction j will equal

$$(4) \quad n_{ij}^{UCP} = m \times \frac{f_i^{UCP} \times N_{ij}^{UCP}}{S_j^{UCP}},$$

where m is the total number of interviews initiated, which is constant across all jurisdictions, except the certainty selections.⁸ This allocation reduces the variation in the sampling weights by ensuring that, *a priori*, all recipients in each BYB date range have an equal probability of being included in the survey sample (Folsom, Potter, and Williams 1987). The n_{ij}^{UCP} individuals will be chosen randomly from the administrative records of jurisdiction j using implicit stratification according to BYB month and study subpopulation, as described above.

4. Stratification of the First-Stage Selection Process

Primary strata for selecting UI jurisdictions in the first stage of the sampling process will be formed to address two important analytic goals of the evaluation: (1) ensuring that the sample includes adequate variability the maximum number of weeks (MNW) of benefits that became available through regular UI, EUC08, and EB combined; and (2) addressing potential bias in the

⁷ Once states have been selected using the joint size measure, the UCP sample will still be allocated across BYB date ranges using equation (2). This continues to ensure that all members of the UCP study population have an equal *a priori* likelihood of being included in the survey sample, reducing the need to apply unequal weights.

⁸ A higher total number of interviewers will be allocated to certainty jurisdictions to account for the fact that they are, in essence, undersampled, relative to the frequency with which they would be selected if the jurisdictions could be drawn with replacement.

survey estimates due to jurisdiction-level nonresponse.⁹ Each of these dimensions of stratification is described in the subsections below; the expected number of jurisdictions to be selected from each primary stratum is discussed in a third subsection.

Stratifying to Ensure Adequate Variability in the MNW Across Jurisdictions. DOL has placed an analytic emphasis on the MNW as a potentially important source driver of differences in outcomes among UI recipients for the UCP study. Consequently, the first-stage selection will also be stratified according to the MNW in each jurisdiction. The distribution of the MNW across jurisdictions suggests defining three strata: (1) 60-79 weeks (12 states; “low”), (2) 86-94 weeks (8 states; “medium”), and (3) 99 or more weeks (30 states and DC; “high”).

Stratifying to Address Possible Jurisdiction-Level Nonreponse. Although the evaluation team will follow Mathematica’s established practices to maximize response rates at every level (see Section B.3), UI jurisdictions may not cooperate with this study’s request for administrative claims data. Based on the experiences of Mathematica staff in conducting a 1990s study of the emergency unemployment compensation (EUC) program (Corson et al. 1999), UI jurisdictions that are experiencing more strain on their UC system due to a worse economy may be less likely to cooperate. This could result in biased survey estimates because differences among states in economic conditions are expected to also affect the individual-level outcomes relevant to the UCP study. To address this potential for nonresponse bias from jurisdictional-level nonresponse, first-stage selection will also be stratified according to the observed increase in the percentage change in UI first claims between calendar year 2007, a period that included the last business cycle peak, to calendar year 2009, a period that covered the trough of the recent recession. This stratification factor was chosen because the percentage change in claims (PCC) can be regarded as a proxy for the recessionary strain on the UC system within a state.¹⁰ Two strata will be formed based on the PCC variable: a “low” stratum containing jurisdictions in which the change in claims ranged from 23 to 74 percent (25 states and DC), and a “high” stratum in which the PCC variable ranged from 82 to 162 (25 states).¹¹

Stratifying on the PCC variable will enable the creation of a randomly-selected reserve sample of UI jurisdictions that has a similar distribution of this measure of recessionary strains as the main sample. In the event that a jurisdiction refuses to provide data after intensive recruitment efforts, an additional randomly-selected jurisdiction from the same primary stratum (defined by the PCC and MNW variables together, as described below) can be released into the sample. Because the random

⁹ The selection of UI jurisdictions will also be implicitly stratified according to geography using three strata based on DOL regions. The first stratum consists of UI jurisdictions in the Northeast, Mid-Atlantic, and South (regions 1, 2, and 3). The second stratum consists largely of states in the Rocky Mountains, the Texarkana area, the Great Plains, and the Midwest (region 5 and most of region 4). The third stratum consists of Pacific and Southwestern states (region 6 and New Mexico). Preliminary simulations of the sampling process suggested that this grouping structure could, on average, achieve a geographic balance across all of the DOL regions. Nonetheless, given that geographic stratification will occur after the sample of jurisdictions is divided into five primary analytic strata (as described in the text), the sampling process is unlikely to ensure an even allocation across regions (or geographic strata) in every sample.

¹⁰ Annual claims data are used, rather than monthly or quarterly data, to avoid having differences across states in the seasonality of unemployment affect the stratification variable.

¹¹ Forming three or more PCC strata is not feasible because, when forming primary strata using both the PCC and MNW variables, over 60 percent of the jurisdictions selected for the analysis would be chosen with certainty, which has negative consequences for the precision and the face validity of the sample.

addition to the sample will have a similar range for the PCC variable, augmenting the sample in this manner should reduce the likelihood that sample estimates are biased by differential nonresponse among states that experienced a certain extent of change in the volume of UI claims.

Sampling Rates by Primary Stratum. Crossing the two dimensions of stratification, the 5 primary jurisdiction-level sampling strata are:

1. Low PCC and low or medium MNW
2. Low PCC and high MNW
3. High PCC and low MNW
4. High PCC and medium MNW
5. High PCC and high MNW

It was necessary to collapse the low- and medium-MNW categories together within the low-PCC stratum because, otherwise, they would only contain four and two jurisdictions, respectively. Even after collapsing the two strata together, the expected number of selections from the resulting primary stratum is very small. As shown in the fourth column of Table B.1, a proportional allocation would have resulted in 0.88 states being drawn, on average, from primary stratum 1 over repeated sampling.

Table B.1. Selection Probabilities Based on the UCP and UCP- COBRA Size Measures

Primary Stratum	Category for PCC Variable	Category for MNW Variable	Proportional Sampling	Oversampling Low- and Medium-MNW Strata		
			Expected Number of Jurisdictions in Sample	Number of Certainty Selections	Number of Random Selections in Main Sample	Number of Jurisdictions in Reserve Sample
1	Low	Low-Medium	0.88	2	1	3
2	Low	High	11.37	3	4	13
3	High	Low	1.14	0	2	6
4	High	Medium	1.94	2	2	2
5	High	High	4.68	2	2	7

Sources: Values for the maximum number of weeks (MNW) variable were calculated using (1) annual UI policy information from the Comparison of State Unemployment Laws series archived by the U.S. Department of Labor, Employment and Training Administration (ETA) <http://www.workforcesecurity.doleta.gov/unemploy/pdf/uilawcompar/> (accessed on 4/12/2011), and (2) weekly trigger notice data for the Extended Benefits and Emergency Unemployment Compensation Act of 2008 programs archived online at http://www.ows.doleta.gov/unemploy/claims_arch.asp (accessed on 4/12/2011). Values of the percentage change in claims (PCC) variable and the size measures used to calculate selection probabilities were constructed based on data on UI first payments and first claims available from ETA online at <http://workforcesecurity.doleta.gov/unemploy/finance.asp> (accessed 01/14/2011).

Table B.1 (continued)

Notes: The figures in the table are based on the assumptions that 20 UI jurisdictions will be selected in the first stage of sampling and in the second stage: (1) 3,000 recipients with BYB dates distributed equally across the four six-month intervals between October 1, 2007, and September 30, 2009, will be selected for the UCP study and (2) 12,000 recipients with BYB dates between February 17, 2009, and May 31, 2010, will be selected for the COBRA subsidy evaluation. Categories for the MNW and PCC variables were defined as described in the text. The expected number of selections with proportional sampling and the number of certainty selections with oversampling of the low- and medium-MNW stratum were calculated using the composite size measure displayed in equation (3).

Given the distribution of the size measure across the five primary strata, it was desirable to oversample in primary strata covering the low- and medium-MNW categories. Taking the oversampling rates into account, the fifth column Table B.1 shows the number of certainty selections in each primary stratum. These nine jurisdictions all would have expected selection frequencies larger than one using the revised sampling rates. The sixth column of the table shows the number of random selections in the main sample for in each stratum. This is equivalent to the number of randomly-selected jurisdictions included in the final sample if there were a 100 percent response rate. The final column of the table displays the number of additional UI jurisdictions in the reserve sample by stratum, which represents the maximum number of additional states that could be released into each stratum in the event of nonresponse in the initial sample.

5. Construction of Weights

Each of the analyses based on the UI recipient survey will use appropriate weights so that the estimates can be generalized to the appropriate population. These weights will be developed using a two-stage process: (1) computation of initial sampling weights; and (2) adjustment of the sampling weights for nonresponse. Each of these steps is discussed below.

Initial Sampling Weights. In the first step, initial sampling weights are computed based on the probability of selection at each of the two stages (UI jurisdictions and individuals within jurisdictions). In the first stage of the sample design, the certainty jurisdictions will have weight of 1 and the randomly selected (noncertainty) jurisdictions will have a sampling that is inversely proportional to the probability of selection. The second-stage weight component will be based on the probability of an individual being selected from the UI claims records. This component will vary within each of the four BYB date range strata described above.

Nonresponse Adjustments. In the second step, the sampling weights are adjusted for nonresponse at both stages. Nonresponse at the jurisdiction level will be handled differently based on whether the jurisdiction is selected with certainty or the jurisdiction is a non-certainty jurisdiction.

A certainty jurisdiction is, by definition, a jurisdiction with a sufficiently large population size that the jurisdiction is unique. Therefore, if a certainty jurisdiction refuses to provide UI administrative claims records for this evaluation, the study population will be redefined to exclude the persons in the noncooperating jurisdiction. Survey estimates will then enable inferences to the population of individuals in the remaining jurisdictions. The redefinition of the population for inferences is a conservative approach since it limits the inferences to a population that had a chance

of inclusion into the study. If a noncertainty jurisdiction refuses to cooperate with a data request, this refusal will be accounted for in the nonresponse adjustment for the individual-level sampling weights.¹²

Individual-level nonresponse adjustments will be made using response propensity modeling and post-stratification. In essentially all surveys, the sampling weights need to be adjusted to account for sample members who cannot be located or who refuse to respond once located. The adjusted weight is the product of the sampling weight and an adjustment factor. The approach to be used in this study to calculate adjustment factors is a generalization of the commonly used method in which “weighting classes” of sample members with similar characteristics are formed and adjustment factors are calculated as the inverse of the weighted response rate in that class. This method produces unbiased estimates of population parameters when the (unobserved) outcomes and characteristics of individuals in the same weighting classes are the same, on average. The natural extension to the weighting class procedure is to use logistic regression with the weighting class definitions used as covariates. The logistic regression approach also has the ability to include both continuous and categorical variables, and standard statistical tests are available to evaluate the selection of variables for the model (Särndal et al. 1992).

Two logistic regression models will be used to calculate nonresponse adjustments. In the first model, the binary dependent variable will be defined according to whether the individual could be located. In the second model, which will be estimated within the sample of individuals who were located, the dependent variable will differentiate between “respondents” and “nonrespondents.” In the UCP study, sample members will be classified as “respondents” if they complete the interview (or if someone does so on their behalf) or if they are determined to be ineligible for the study (for example, if they are deceased). Based on these logistic models, the inverse of the propensity scores will be used as adjustment factors. The adjusted weight for each sample case is the product of the initial sampling weight and the two adjustment factors.

Each logistic nonresponse model will be fitted by first identifying a pool of covariates to work from using stepwise regression, then assessing candidate models using various measures of goodness of fit and predictive ability. The covariates will include factors or attributes that can be obtained from administrative data and (1) which are likely to be associated with differences in the likelihood that a sample member is located and interviewed and (2) have been shown by previous research (Corson et al. 1999; Needels et al. 2000) to be related to the outcomes of interest for this study among UI recipients. Specific examples include:

- Pre-claim earnings, occupation, and industry
- Reason for separation from pre-claim job
- Age
- Gender
- Race and ethnicity
- Geographic location

¹² Additional adjustments may be made based on the findings of the nonresponse analysis described in Section B.3.

A chi-squared automatic interaction detector (CHAID) will be used to refine the list of candidate independent variables and identify interactions among them.¹³ The CHAID procedure iteratively segments a data set into mutually exclusive subgroups that share similar characteristics based on their effect on nominal or ordinal dependent variables. It automatically checks all variables in the data set and creates a hierarchy that shows all statistically significant subgroups. The algorithm finds splits in the population, which are as different as possible based on a chi-square statistic. It is a forward stepwise procedure, and it finds the most diverse subgrouping, and then each of these subgroups is further split into more diverse sub-subgroups. Sample size limitations are set to avoid generating cells with small counts. The algorithm stops when splits no longer are significant; that is, the group is homogeneous with respect to variables not yet used or when the cells contain too few cases. The CHAID procedure results in a tree that identifies the set of variables and interactions among the variables that have an association with the ability to locate a sample member and the propensity of a located sample member to be a respondent (eligible or ineligible).

The variables and interactions identified using CHAID then will be processed using forward and backward stepwise regression (using SAS Logistic procedure with weights normalized to the sample size) to further refine the candidate variables and interaction terms. After identifying a smaller pool of main effects and interactions for potential inclusion in the final model, a set of models will be evaluated to determine the final model. Because the SAS stepwise logistic procedures do not incorporate the sampling design, the final selection of the covariates will be accomplished using the logistic regression procedure in SUDAAN (Research Triangle Institute 2004).

After the nonresponse adjusted weights are computed, survey estimates of the number of UI recipients with first payments in each BYB date range will be post-stratified to the national counts available from ETA. In some situations, the post-stratification factors or nonresponse adjustment factors can introduce excessive variation in the sampling weights, which can reduce the precision of survey estimates. Consequently, extreme weights might be trimmed using one of the methods by Potter (1990, 1993) that reduces sampling variation while minimizing the potential for bias caused by trimming. The weights again will be post-stratified to population counts after the weight trimming. Because the sampling design will result in nearly equal weights in the BYB date range strata, there is likely to be little or no weight trimming (Potter et al. 1998).

b. Survey of UI Administrators

The sample for the survey of UI administrators will be the 51 major UI jurisdictions (the 50 states plus the District of Columbia). There will be no subsampling for survey administration. The key outcomes of interest for the regression analysis of state decision-making are indicators for whether the jurisdiction adopted each of the five UI modernization provisions of interest and the total unemployment rate (TUR) trigger for EB benefits. Responses to the survey questionnaire will be used to create explanatory variables to be included in separate regressions for each outcome.

¹³ CHAID is normally attributed to Kass (1980) and Biggs et al. (1991), and its application in SPSS is described in Magidson (1993). Decisions about variables and interactions will be based on statistical tests with the significance level (alpha level) set to 0.30. The test size of 0.30 is used instead of the standard 0.05 because the purpose of the model is to improve the estimation of the propensity score and not to identify statistically significant factors related to response.

c. Site Visit Data Collection

Understanding states' experiences in implementing changes to their procedures as a result of ARRA, the challenges they faced, and strategies they used to overcome those challenges is useful for shedding light on likely responses and successful strategies if similar policies were to be implemented at a later time. In addition, these implementation experiences provide context to interpret the effects of the UC-related provisions of ARRA on UC claimants. The main source of information on states' implementation experiences will be site visits to 20 states to gather information from multiple respondents about their experiences implementing the UC-related provisions of ARRA.

In order to fully represent this diversity, the sampling plan requires selecting states purposively. As a first step, the study team has identified the following variables to capture the desired diversity across the 20 chosen states:

- Adoption of optional provisions: TUR trigger for EB, alternate base period (ABP), part-time work provision, compelling family reasons provision, dependents' allowance provision, and training provision
- TUR as of a particular date
- Size of state UI program, as measured by total UI first payments from 2008–2010
- Geographic location

Including 20 states in this data collection effort will allow for learning about a broad range of approaches and experiences, including states that made significant changes to qualify for the incentive funds, ones that qualified for the funds but did not need to make significant changes, and ones that did not apply for incentive funds.

Using data collected from public sources as well as information collected through the survey of UI administrators, the study team will construct and fill a table with these variables for all states (Table B.2). First, the team will identify whether states already had a provision, newly adopted it, or did not adopt it (and which provisions were adopted). This information will be recorded as columns in the table. We will also categorize the state TUR as of specified dates, the categories for which will be shown as rows in the table. Then, within each cell, states will be sorted by the number of UI first payment recipients. Finally, within these cells, the study team will purposively choose states to reflect a range of experiences. Given the large number of characteristics proposed, the study team will apply the geographic criterion after completing the selection process to ensure that the sample of 20 states is geographically diverse. To the extent possible, states' responses to a question on the survey of UI administrators about whether the decision to adopt was characterized by intense debate will also factor into the selection.

Data on implementation experiences gathered through the site visits will be analyzed primarily using qualitative methods. When states' experiences are quantifiable, they will be tabulated, and narratives of common themes and patterns in states' implementation experiences will be constructed. No statistical inference will be used in this analysis.

Table B.2. States' Characteristics by Receipt of UC Modernization Funds Received (as of December 2010)

	Size ^b	Received Full Modernization Funds		Received ABP Portion Only		No Provisions
		Already Had ABP or Other Provision	Newly Adopted All 3 Provisions	Already Had ABP	Adopted ABP	
TUR ^a						
Less than 6 percent	Small	NH	NE; SD	VT		ND
6.0 to less than 8.5 percent	Small	AK; <i>HI</i> ; ME; NM	DE <i>MT; OK</i>		<i>UT</i> ^c	<i>WY</i>
	Medium		AR; CO; IA; KS; MD; MN	VA		LA
	Large	MA; NY; WI				PA; TX ^c
8.5 percent or greater	Small	DC; RI	ID		WV	<i>MS</i>
	Medium	CT; NV	OR; SC; TN			AL; AZ KY; MO ^d
	Large	IL; NC; NJ	GA	MI ^c ; OH; WA		CA ^d FL ^c ; IN

Sources: State TUR is taken from Trigger Notice No. 2010-49; modernization funds received and ARRA-specified provisions adopted are from state certifications for modernization incentive funds, <http://www.doleta.gov/recovery/#PressReleases>; size of UI system from monthly reports found at <http://workforcesecurity.doleta.gov/unemploy/claimssum.asp>.

Notes: Italics: TUR trigger not adopted as of December 19, 2010.

Bold: did not trigger on to EUC08 tier 4 as of December 19, 2010.

ABP = alternate base period; TUR = total unemployment rate; UI = unemployment insurance.

^aReflects average seasonally adjusted TUR for three-month period ending November 2010.

^bSize reflects the number of UI first payments made between January 2008 and October 2010. During this period, small states made fewer than 250,000 UI first payments; medium states made between 250,000 and 749,999 UI first payments; and large states made 750,000 or more UI first payments.

^cLegislation to adopt all or some of the provisions did not get through the state legislature.

^dLegislation to adopt all or some of the provisions was passed but did not meet ARRA requirements; application for incentive funds was not approved.

2. Analysis Methods and Degree of Accuracy

Four subsections present information about the methods used as part of the evaluation of the UC provisions of ARRA to analyze (1) state decision making, (2) implementation of the UC-related provisions of ARRA on states, (3) descriptive information about the characteristics of UC recipients, and (4) estimation the impacts of UC-related provisions of ARRA on UC recipients.¹⁴ A fifth subsection presents information about the precision of the estimates based on the UC recipient survey.

¹⁴ As explained earlier, the study has five topic areas. The fifth topic area, which pertains to the influence of the UC provisions of ARRA on macroeconomic issues and state UI trust funds, will not use data that are part of this clearance request; the data to be used are publicly available.

a. Analysis of State Decision Making

The study of state decision making will be primarily informed by data collected from the survey of UI administrators. This survey will be sent to the UI administrators of all 50 states and the District of Columbia. Since the format of the survey includes primarily closed-ended questions, the data collected will support a descriptive analysis, including a quantitative regression analysis.

The analysis will first document the adoption decisions of each state for each provision. Second, using publicly available information on states, the analysis will summarize the economic and political characteristics of states, such as the unemployment rate, the UI reciprocity rate, and the political party controlling the state legislature and governorship at the time ARRA was introduced. To detect patterns in the decision-making process, the study team will group states into categories based on the status of various provisions: whether they already implemented a provision, modified an existing provision, adopted a new provision, or did not adopt a given provision.

The third part of the descriptive analysis will directly examine the reasons why state decision makers did or did not adopt the provisions. To complete this analysis, the study team will use data gathered from the survey of UI administrators to tabulate states' responses to closed-ended questions about the key factors for and against adoption and the nature of the discussion surrounding whether to adopt the provisions (such as whether there was intense debate). Then, the characteristics of states that shared similar adoption processes will be described in order to discern any trends or common characteristics.

The regression analysis will draw on publicly available data and responses to the survey of UI administrators to determine whether there are statistically meaningful factors that predict states' adoption decisions about the TUR trigger and the five modernization provisions. States will be the unit of observation and adoption of a given provision will be a binary outcome variable (with separate models for each of the six provisions being investigated). The analysis will employ explanatory variables, measured prior to ARRA, of four broad types:

1. State labor market variables such as the TUR or Insured Unemployment Rate (IUR) and a measure of unionization
2. UI statutes such as the base period earnings requirement, the statutory benefit replacement rate, and maximum weekly benefits as a percentage of average weekly wages
3. UI performance variables such as the UI reciprocity rate and a measure of UI trust fund reserve adequacy
4. Variables that reflect the political situation in the state such as the political party of the governor and the two houses of the state legislature

The regressions will take the general form:

$$Adoption_i = f(\mathbf{X}_i) + \varepsilon_i,$$

where $Adoption_i$ is a binary variable taking a value of one if state i ever adopted the ARRA-specified provision, and a value of zero if it did not; \mathbf{X}_i includes some or all of the variables mentioned in bullets 1-4 above, and ε_i is an error term. Because there is likely to be a collinear

relationship among the explanatory variables, the study team will employ methods, such as a stepwise approach where one set of explanatory variables is added, followed by another, to determine which of the variables are the best predictors of provision adoption.

Furthermore, it is possible that states already had similar UC provisions in place prior to ARRA. To learn more about the relationship between political, economic, and other characteristics of states, the analysis will include an examination of how states' characteristics are associated with adoption of a specific UC provision before ARRA, as well as how the characteristics are associated with adoption of the provision after ARRA. In addition, states that have a provision in place (such as an ABP or a dependents' allowance) may have been more likely to fully adopt ARRA-specified provisions than states that had no related provisions in place. To determine whether this was the case, a set of ordered probit models will be estimated. These models have the same general form as the regression equation above except that the dependent variable could take a value of two if the state newly adopted the provision, a value of one if the state already had a similar provision in place, and a value of zero if the state did not adopt the provision. By using these approaches, the analysis has the potential to provide insights about the adoption of provisions both prior to and after ARRA.

For both types of regression-based analyses, standard inferential techniques will be used to determine the statistical significance of explanatory variables. In particular, the study team will use two-tailed tests with the level of significance set to 5 percent. With only 51 observations in each estimated regression equation at most, and possibly considerably fewer, it may be difficult to detect statistically significant relationships between the explanatory variables and the adoption decision. However, as part of the examination of the validity of the models, the analysis will include checks of the robustness of the results to alternate specifications of the models; the sensitivity of the results will be highlighted, as needed, when conclusions are presented. The tabular analyses outlined above also will supplement the regression analyses and help the evaluation team further flesh out the relationships among the characteristics of states and their decisions whether to adopt ARRA-specified UC provisions.

As noted further in Section B.3, the study team anticipates that the survey of UI administrators will be completed by all 51 targeted respondents and has identified methods to support attaining this response rate. However, in the case of nonresponse by one or more states, the study team will assess the extent to which survey nonrespondents differed from respondents by comparing the economic and political characteristics of the two groups of states, as well as their adoption experiences (which will be known to some extent from publicly available sources even if they do not respond to the survey). The small number of total respondents and the very small anticipated number of nonrespondents will most likely preclude any statistical tests of differences in the characteristics of nonrespondents and respondents. However, simply noting whether such differences appear to exist will be helpful in determining the extent to which the groups differed. This information will be incorporated into the discussion of the results of the descriptive and regression analyses.

The analysis of state decision making will include a final component consisting of a detailed qualitative analysis of the decision-making processes in the 20 states visited. The interview protocols for the site visits will include modules of questions on the decision-making process; only respondents with knowledge of the process will be asked these questions. These respondents will include the UI administrator, key legislators and lobbyists, and members of the state advisory council on UC. As described in further detail in the next section, the interviews will be coded using a qualitative analysis software package and analyzed in much the same way as the responses to the survey of UI administrators, described above.

b. Analysis of the Implementation of the UC-Related Provisions of ARRA on States

Part of the evaluation is an analysis that will be used to document the states' experiences in implementing the ARRA-related UC program provisions. Because UC programs differ across states, and because they operated in very different environments, there is no single, precise, and uniform implementation experience at states across the country. In recognition of this, the analysis will identify both themes that span the states and distinctive features or patterns that occur in only a subset of states.

Site visits in 20 states will serve as the main sources of data for the implementation analysis. The visits will take place following the completion of the fielding of the survey of UI Administrators. Data collection and reporting procedures will ensure that the study will capture the diversity of states' experiences and the perspectives of multiple respondents in each state.

Site visitors will use a write-up template to create a narrative of the interviews conducted as part of the site visits; the write-ups will describe how states implemented the ARRA-related UC provisions, the challenges they faced, and the effects of enacting the provision(s). Because analyzing data from multiple respondents can be complicated, the study team will sort and code the site visit narratives to ensure that the analysis includes all perspectives and that the team can count and report the number of states with similar experiences. The 20 narrative reports will be compiled in a database using Atlas.ti qualitative analysis software for coding (ATLAS.ti 2011). Atlas.ti enables the research team to use a structured coding system for organizing and categorizing data, entering the data into a database according to the coding scheme, and retrieving data that are linked to key research questions. Researchers will use the coded data to tabulate common experiences across the states and look for patterns to help facilitate the development of hypotheses.

Using the coded site visit data, the study team will conduct a cross-state analysis of states' implementation of the ARRA provisions and the factors that shaped their experiences. These analyses will use the state as the unit of analysis, and will primarily tabulate states' experiences (for example, 5 of the 15 states that implemented an ABP faced significant challenges in modifying their data systems).

An important part of the implementation study will be ensuring the accuracy and reliability of both the data and the conclusions derived through analysis of the data. As described in more detail in Section B.3, strategies to ensure that the data are reliable and as complete as possible include using a flexible approach to schedule visits and assuring respondents that the information they provide will remain private. Furthermore, using structured, pre-determined protocols to collect the data and thoroughly training the site visitors will help achieve a high degree of accuracy in the data. Because most questions will be asked of more than one respondent during a visit, the analysis will allow for comparisons and triangulation of the data so that discrepancies among different respondents can be interpreted.

c. Descriptive Analysis of the Characteristics of UC Recipients

Data from the UI recipient survey will be used to describing the characteristics of a study population consisting of UI recipients with BYB dates between October 1, 2007, and September 30, 2009. The EUC08 program and complete federal funding of EB were both intended to extend the duration of unemployment benefits, providing additional income support to workers who were experiencing long spells of unemployment. The appropriateness of these benefits policies depends, in part, on the types of people who received benefits from the programs they established.

To shed light on this issue, this study will describe the characteristics of recipients of extended benefits and compare them with those of other groups of unemployed workers, particularly recipients of regular UI benefits. The descriptive analysis will also consider how the characteristics of recipients of regular UI and extended benefits differed according to the duration of benefit receipt, the UI policies enacted by the jurisdictions in the sample, and economic conditions that varied across time and between UI jurisdiction. Comparisons will also be made between the recipients of extended benefits during the 2007–2009 recession to longer-term recipients of UC benefits during earlier recessionary and nonrecessionary periods using data from four studies previously conducted by Mathematica for DOL (Corson et al. 1977; Corson, Grossman, and Nicholson 1986; Corson, Needels, and Nicholson 1999; and Needels, Corson, and Nicholson 2002).

Furthermore, the study will include an in-depth descriptive analysis of the employment and training outcomes of recipients of the two types of extended benefits and other unemployed workers in order to provide a departure point for the analysis of program impacts (discussed in the next subsection). The analysis will examine (among other topics) how long recipients remained unemployed, how long they collected unemployment benefits, the nature of the work search, education, and training activities that recipients engaged in while unemployed, and the characteristics of the first post-UI job among individuals who became reemployed. This descriptive analysis will also consider differences across subgroups of UC recipients in the receipt of other forms of government assistance (such as those that provide benefits to low-income households), as well as in their income levels prior to receipt of UI benefits and any financial hardships they experienced during the unemployment spell. Finally, the descriptive analysis will characterize the distribution of the dollar value of UC benefits across recipients to provide a better understanding of intergroup differentials in UC payments received and how these differentials related to UC policies.

Methods for Calculating Point Estimates. Many of the descriptive analyses will be based on simple weighted summary statistics.¹⁵ For example, comparisons between subgroups may be based on the difference in means or proportions. When considering employment and benefit durations, the analysis will rely on the conditional probability of reemployment between two time periods among recipients whose outcomes are observed in both time periods. This conditional probability, referred to as the Kaplan-Meier hazard rate, will be used as a summary measure to avoid the biases from censoring that would occur because some people will still be jobless at the time of the study's follow-up interview. More sophisticated regression-based models, such as those described in the following subsection about the impact analyses, may also be used for descriptive purposes because they can better isolate the independent relationship between a single attribute and recipient outcomes. All of the descriptive estimates will be calculated using analytic weights that account for the survey sampling methodology, including a nonresponse adjustment.

Variance Estimation for Descriptive Measures. Test of significance for point estimates and contrasts calculated in the descriptive analysis will be based on variance estimates that explicitly account for the complex survey design, for example, clustering, stratification, and weighting. These design-based variances will be estimated using Taylor linearization (see, Binder 1983 and Sections 5.5 through 5.10 of Särndal et al. 1992) as implemented in SUDAAN, SAS, or Stata. (In Särndal et al. [1992], equations 5.5.7 and 5.5.8 present the basic equations for the first-order Taylor series

¹⁵ All survey estimates are design-based and will be computed using the design-based sampling weights adjusted for nonresponse.

approximation; the application of the Taylor series approximation for variance estimation of ratios is given in Section 5.6, for means in Section 5.7, and for regression coefficients in Section 5.10.) A finite population correction will not be made at either the individual level or jurisdiction level so that the study will have some capacity to generalize inference based on the results beyond the study population.

d. Impact Analyses

The evaluation will apply rigorous quasi-experimental methods to data from the UI recipient survey in order to analyze the effects of changes to UC provisions under ARRA on the following major categories of outcomes:

- Duration of the initial unemployment spell
- Duration of UC benefit receipt
- Earnings upon reemployment
- Measures of financial hardship
- Work search intensity near the start of the benefit spell
- Likelihood of participating in education or training programs

The statistical methods used in the analysis will rely on variation across UI jurisdictions, among claimants, or over time to estimate the impacts of the policies on recipients' outcomes.¹⁶ Because policies and program rules are often changed in response to evolving economic conditions, causal impacts will be identified based on sharp changes in behavior attributable to policy changes.

Table B.3 summarizes the sources of variation for each of the UC policies considered in the impact analysis using data from the UI recipient survey.¹⁷ As seen in the table, the specific methods used to estimate the impact of a given policy change will depend on the nature of the variation in that policy. Changes that occurred across the whole nation at the same time—for example availability of EUC08 Tier 1 benefits—must be analyzed using an interrupted time series (ITS) design. Policy changes that were staggered across UI jurisdictions or that occurred in some jurisdictions but not others—for example, availability of additional weeks of benefits through EUC08 Tier 4 or EB—may be analyzed using more rigorous methods such as differences in differences (DD) and regression discontinuity (RD).

¹⁶ Additional analyses using administrative data will consider how ARRA-based changes to UC policy may have affected the composition of the recipient population as well as how eligibility for UI under one of the modernization provisions affected the outcomes for unemployed workers who might not otherwise have qualified for UI.

¹⁷ The table does not include the UI modernization provision setting a floor on the increment to the WBA for recipients with dependents. Although the availability of dependents' allowances will be controlled for in the analysis, it is not likely to be possible to draw meaningful conclusions about the impacts of this provision on recipients' outcomes. The reason is that only three states (Illinois, Tennessee, and Rhode Island) implemented new dependents' allowance provisions after ARRA was implemented. Furthermore, only one (Illinois) has a high probability of being included in our sample of states.

Table B.3. Estimation of Impacts

Unemployment Compensation Policy	Source of Variation	Method Used
EUC08 Tiers 1 and 3	UI jurisdiction benefit formulas; timing of availability	ITS
EUC08 Tier 2 and 4; Extended Benefits	UI jurisdiction benefit formulas; jurisdiction-by-time variation in availability; IUR/TUR triggers	DD/RD
Training-specific 26-week extension	Jurisdiction-by-time variation in availability	DD
Federal Additional Compensation	Timing of availability	ITS
Tax exemption on first \$2,400 of UI benefits		

DD = differences in differences; EUC08 = Emergency Unemployment Compensation Act of 2008; ITS = interrupted time series; IUR = insured unemployment rate; RD = regression discontinuity; TUR = total unemployment rate; UI = unemployment insurance.

Each of the methods used to estimate impacts is discussed in detail in the first subsection below. The second subsection describes additional considerations for accounting for censoring when analyzing recipients' duration-dependent outcomes. The final subsection explains the approach that will be used to test the statistical significance of the impact estimates.

Methods for Estimating Impacts. The basic statistical approach to estimating policy impacts on recipient outcomes relies on a linear equation of the form:

$$(5) \quad y_{ist} = \beta' p_{st} + \eta'_{st} x_{ist} + \theta' z_{st} + \alpha_s + \varepsilon_{ist},$$

which will be estimated using weighted least squares.¹⁸ As with the descriptive analyses, the analytic weights will be based on the survey sampling methodology, including any adjustments made for nonresponse, so that regression analyses using the sample of UI recipients produce representative estimates for the nation as a whole.

The outcome variable is y_{ist} , where s denotes the UI jurisdiction in which the individual i receives benefits, and t denotes his or her "cohort," defined by the month in which UI benefits were first received. To the extent possible, the analysis will focus on outcomes that have been measured at some common interval after job loss (or after the initial UI claim), such as 12 months and 24 months. Setting a common time of observation ensures that individuals are being compared at similar different points in their unemployment spell.¹⁹ The main exception is that data may be

¹⁸ For ease of exposition, the outcome variable is assumed to be continuous. When considering binary outcomes, equation (5) could be re-specified as a nonlinear probit or logit model. However, a regression coefficient from a linear probability model often provides a reasonable approximation to the marginal effect of a variable that would be obtained from a nonlinear binary response model (Wooldridge 2002). Because of its substantial advantages for computation and interpretation, the linear model will be used if the regression coefficients are similar to the marginal effects obtained from the nonlinear model.

¹⁹ The analyses the impact of UI policies on postunemployment outcomes might exclude from the sample individuals who started receiving benefits more than one month after losing their jobs. The main reason for this restriction would be to ensure that comparisons are based on individuals who faced similar economic conditions during their unemployment spell and to whom similar UI policies were applicable. Simultaneously accounting for both the date of job loss and the date of entry into the UI system is likely to result in too many control variables, given the fixed-effects approach used in many of the analyses. Preliminary analyses will investigate the potential implications of this restriction by, for example, determining what portion of recipients is affected.

pooled from multiple times of observation for the same cohort when analyzing unemployment duration, as discussed in greater detail below.

The term p_{st} in equation (5) describes the set of UC policies in effect in jurisdiction s that may affect the outcomes of cohort t . These policy variables would include, for example, whether a given tier of EUC08 or EB benefits was available immediately when each recipient would have exhausted benefits under the next-lowest tier with continuous collection of the full weekly benefit amount (WBA). An alternative approach is to interact the policy change variables with an individual's baseline MNW of benefits, as defined by jurisdiction-specific UI policies. This could better take into account the fact that individuals with a higher baseline MNW (up to 26 weeks) qualify for more additional weeks of benefits when new tiers of EUC08 or EB benefits become available. Policy variables may also be interacted with the WBA, which affects the financial value of the additional weeks of available benefits. In addition, p_{st} may measure the status of policies at the start of the spell (time t) as well as changes to policies affecting an individual's potential benefits that occurred after time t . Finally, the policy variables might include the fraction of time (from t to the time of observation) that a 26-week training extension was in place in jurisdiction s .

In general, the policy variables will be specified so that estimated impacts (captured by β) are based the average response of individuals to changes in the policies affecting the benefits for which they were *potentially* eligible. This approach is sometimes referred to as an "intention-to-treat" (ITT) framework and will be used to avoid the bias that would result from focusing only on individuals who actually responded to a policy. For example, individuals who actually made use of extended benefits are likely to differ substantially from individuals who did not claim the additional benefits made available to them. Most problematically, individuals who did not make use of the extra benefits are much more likely to have found a job before exhausting regular UI, whereas individuals who moved onto EUC08 or EB remained, by definition, unemployed.²⁰ By using data on all individuals who were potentially affected by a policy, the ITT framework will produce estimates that do not suffer from this form of choice-based bias and likely have greater salience for policymakers interested in the overall effects of UI policies.

The regression includes a set of individual-level control variables, x_{ist} such as base period earnings, age, race and ethnicity, gender, marital status, education, family size, and occupation. These characteristics will all be measured prior to the claim to avoid confounding the estimated policy effect.²¹ Equation (5) also has controls for time- and jurisdiction-specific economic conditions, z_{st} , which may include the unemployment rate, income per capita, and industrial composition measured

²⁰ Propensity score matching cannot be used in this setting to reduce bias because the one reason that individuals would not claim extra weeks of benefits under EUC08 is that they found jobs before exhausting their basic UI entitlement. This violates the central assumption of propensity score matching described in Rosenbaum and Rubin (1983) because the likelihood of being assigned to the "treatment" of receiving EUC08 benefits is explicitly contingent on labor market outcomes.

²¹ Because the analysis is limited to recipients, x_{ist} may additionally contain a "Heckman selection correction" term (Heckman 1979) calculated from an auxiliary analysis of the likelihood that an individual will receive UI based on administrative data. This extra term is intended to adjust for bias resulting from compositional changes in the recipient population induced by policy changes. Specifically, the term accounts for unobservable factors that might affect both the outcome of interest and the likelihood that an unemployed worker receives benefits

just before the start of the claim and at various points between time t and the time of the UI recipient survey. Finally, the statistical model includes jurisdiction intercepts, α_s , which are specified as fixed effects, rather than random effects. This approach results in stronger internal consistency for estimating causal effects, since random intercepts are constructed to be, by assumption, uncorrelated with the explanatory variables. By contrast, jurisdiction fixed effects explicitly account for persistent differences in unobserved jurisdiction-specific factors that may be correlated with the decision to implement specific ARRA-based changes to UI policies, such as a 26-week training extension or a TUR trigger for EB.

The main concern for treating the estimated relationship between policies and outcomes ($\hat{\beta}$) as measuring causal impacts is that policy changes could be correlated with the unmodeled determinants of outcomes of individuals, ε_{ist} . Most problematically, policy changes may occur at different times in different jurisdictions in response to unmeasured changes in the labor market, which would result in biased impact estimates.

The ITS, DD, and RD designs all refine the basic specification in equation (5) to reduce this potential bias. Each is described below. When using each of these specifications, sensitivity checks will be conducted to assess the robustness of the resulting impact estimates. Such checks may include testing for a change in outcomes that precedes a change in policy or determine whether policies have a significant association with outcomes that they should not affect—both might suggest that the specification is not effectively isolating causal impacts of the policies of interests.

ITS Design. The ITS design modifies equation (5) so that the statistical framework accounts for preexisting trends in each jurisdiction:

$$(6) \quad y_{ist} = \beta' p_{st} + \eta'_{st} x_{ist} + \theta' z_{st} + \alpha_s + \gamma'_s t + \varepsilon_{ist}.$$

The trend variables ($\gamma'_s t$) account for a preexisting pattern of linear change in the unmeasured jurisdiction- and time-specific characteristics. The ITS framework assumes that the added trend variables sufficiently account for changing jurisdiction-level unobserved factors that simultaneously affect outcomes and policy changes, so that the remaining variation in policy may be regarded as random. However, the main limitation of the ITS design is that it cannot account for any unobserved changes that have a similar effect on all members of a UI recipient cohort (indexed by i). Thus the estimated policy effects on earnings could be potentially confounded with a nationwide shift in unmeasured economic conditions that occurred at the same time a national UC policy was enacted or changed.

DD Design. The DD design strengthens the estimation framework further by adding time fixed effects:

$$(7) \quad y_{ist} = \beta' p_{st} + \eta'_{st} x_{ist} + \theta' z_{st} + \alpha_s + \gamma'_s t + \mu_t + \varepsilon_{ist}.$$

Given that outcomes generally will be estimated at a single common time for each UI recipient cohort, time fixed effects (μ_t) are mathematically equivalent to cohort fixed effects and account for unmeasured characteristics of a cohort or unmeasured economic shocks faced by the cohort between job loss and the time of the UI recipient survey.

Jurisdiction and time fixed effects together form the basis of the DD design. Jurisdiction fixed effects will account for the ongoing contribution of baseline differences between jurisdictions. Jurisdictions that do not experience policy changes are used to estimate a common time fixed effect in each period, which is assumed to be the counterfactual change that “treatment” jurisdictions (those that had a policy change) would experience if they had not made any changes to UI policies. Netting out baseline differences and the period-specific differences experienced by comparison jurisdictions gives the DD estimate of the effect of the policy.

The DD design requires that policies changed at different times in different jurisdictions; otherwise, the time fixed effects would perfectly explain the status of the policy. Thus, this design may not be applied when analyzing the FAC, tax exemptions on UI benefits, and benefits from EUC08 Tiers 1 and 3.²² Those policies must, then, be analyzed using an ITS design.²³

RD Design. This analytical approach can be applied when a rule based on a continuous numerical variable (a “forcing” variable) is used to determine the status of a policy. States that fall above a cutoff score of this forcing variable become eligible for a specific policy or benefit, while states below the score remain ineligible. Thus, an RD design may potentially be applied to estimate the impacts of EUC08 Tiers 2 and 4 and EB, all of which have been contingent on IUR or TUR triggers. When using an RD design, the effect of a policy change is estimated near the threshold value of the forcing variable. The regression framework is modified to include a “forcing function,” $g(z_{st})$, which estimates the underlying relationship between the outcome and the forcing variable, denoted as z_{st} :

$$(8) \quad y_{ist} = \beta' p_{st} + \eta'_{st} x_{ist} + \theta' z_{st} + \alpha_s + \mu_t + g(z_{st}) + \varepsilon_{ist}.$$

In a unified regression that simultaneously considers the availability of EUC08 Tiers 2 and 4, as well as EB, z_{st} will include (1) the IUR, (2) the TUR, and (3) the ratio of each of these measures to their associated values one and two years before.²⁴ Equation (8) continues to include fixed effects to address any pervasive differences in the volatility of the unemployment rate trigger variables across jurisdictions and over time as well as differences in the propensity to adopt alternative triggers for EB. A common forcing function can be fit using all of the data points or the function may be fit separately for the states in which a given set of benefits became available and for the states that never triggered onto that tier or type of benefits. Differences in the actual availability of benefits across states near a trigger unemployment rate may be considered to be functionally random, once the forcing variable has been properly controlled for in the regression.

²² The legislation implementing EUC08 Tier 3 indicated that such benefits would go into effect only when the unemployment rate cleared a trigger value. However, the 48 UI jurisdictions that had triggered onto Tier 3 benefits by April 2011 did so immediately after the legislation was passed in November 2009 and have remained eligible for those benefits through the present. Thus, for analytic purposes, Tier 3 benefits must, in essence, be considered a one-time nationwide change.

²³ When the impacts of other policy changes are estimated through use of the DD method, the effects of the FAC, tax exemptions, and EUC08 Tiers 1 and 3 will be implicitly controlled for by time fixed effects.

²⁴ If applicable, measures of the IUR and TUR relative to their values during the three most recent prior years will be constructed for jurisdictions that adopted the three-year “look-back” provision that came into effect in late 2010.

Although RD is viewed as providing strong evidence for quasi-experimental impact estimates, if the assumptions that underlie it hold (Cook and Wong 2008), it may be not be feasible to implement this approach in practice for three reasons. First, sensitivity analyses might suggest that key assumptions of RD do not hold. For example, the distribution of observable characteristics may be very different in the regions above and below the threshold of the forcing variable or that the impact estimates are affected by the bandwidth of data around the cutpoint that is used in the analysis. Second, as described above, RD requires a forcing variable that determines the status of policy. As a result, this methodology may only be applied to selected tiers of EUC08 and to EB. Third, RD greatly reduces the statistical power to detect significant effects because much of the variability in the availability of benefits will be explained by the forcing function. With a clustered sampling scheme, an evaluation that uses an RD design typically needs a sample three to four times as large as an evaluation of the same intervention that uses random assignment (Schochet 2009). This problem is amplified in the case of EB because multiple trigger rules might be in effect. In this case, it also may not be possible to reliably control for all of the trigger variables at once because they are likely to be highly correlated with one another. This would substantially weaken the validity of the RD design. Hence, if RD is not feasible, the DD framework may be the primary framework for estimating impacts.

Accounting for Censored Data. Some of the UI recipients in the sample will not have returned to work at the time of their interview. For these sample members, the duration of unemployment will be censored: neither this duration nor the postunemployment earnings will be observed. When analyzing the duration of unemployment, censoring implies that the observed length of an individual's jobless spell at the time of the survey will underestimate the true length of the jobless spell. This will result in biased regression estimates of the impacts of changes due to UC policies, particularly if the duration of unemployment is affected by unobserved individual-level characteristics.

To address censoring, inferences about unemployment durations will be made by analyzing the probability of reemployment, conditional on an individual not having already become employed. This conditional reemployment probability is referred to as the "hazard rate" and effectively excludes individuals whose spells have been censored at the time of measurement. There are several approaches to estimating effects of changes to UI policy on the hazard of reemployment. One extensively used approach involves estimating parametric models of reemployment on the basis of specific assumptions about the distribution of the hazard (see, for example, Newton and Rosen 1979; Katz and Ochs 1980; and Kruse 1988). However, economic theory does not suggest an appropriate distribution for the hazard, and the magnitudes of estimates made using parametric approaches are often quite sensitive to the chosen distribution (Moffitt 1985). Thus, the analysis will consider semiparametric approaches (Meyer 1990) or the repeated-outcome method described by Kalbfleisch and Prentice (2002). The repeated-outcome method is particularly useful because a linear probability model (LPM) may be applied to analyze the data. As noted above, the LPM typically provides close approximations to the marginal effects of changes in policy, while requiring fewer computational resources. An additional benefit of using an LPM is that it allows individual heterogeneity to be taken into account by specifying individual-level fixed effects.

Two approaches may be used to account for censoring when analyzing postunemployment earnings. The first approach simply sets earnings to zero for individuals who had not become reemployed by the time of the UI recipient survey, which maintains the spirit of the intention-to-treat analysis for policy variables. However, reemployment earnings among job finders are also of substantive interest. With the exception of McCall and Chi (2008), very little work has examined reemployment earnings while accounting for differential selection into employment. Consequently,

the second approach uses a two-step procedure whereby the first step estimates a probit model to predict the likelihood of reemployment by the time of the survey. Applying the estimated coefficients using the properties of the standard normal distribution results in a Heckman correction term that can be added to the DD framework in equation (7) in the second step of the estimation process. The correction term controls for compositional changes in the pool of individuals who re-enter employment by the time of observation, reducing potential for bias when estimating the impact of policy changes (Heckman 1979).

Variance Estimation for Impact Estimates. As with the descriptive point estimates, variances for the estimated impact parameters (denoted as $\hat{\beta}$ above) can be estimated using Taylor linearization in SUDAAN, SAS, or Stata. (See the references provided previously on the use of the Taylor series approximation for variance estimation.) Such variance estimates will take into account variation in $\hat{\beta}$ arising from the design of the survey.

In certain settings, empirical analyses of labor market data have found that design-based variance estimates may not fully account for serial correlation within clusters (primary sampling units) over time when calculating DD impact estimates (Bertrand et al. 2004 and Cameron et al. 2008). Consequently, the evaluation team will explore the feasibility of applying cluster-robust corrections (Bertrand et al. 2004; Froot 1989) or cluster bootstrap methods (Cameron et al. 2008) when conducting statistical inference on the impact estimates. As with the descriptive estimates, finite population corrections will not be used when calculating variances for the impact estimates because one of the goals of the study is to add more rigorous evidence to the existing knowledge base that considers how extended benefits programs might affect the outcomes of UI recipients.

e. Precision of Estimates from the UI Recipient Survey

This subsection considers the precision of estimates computed using data from the UI recipient survey and provides illustrative calculations for the minimum statistically detectable differences that are expected when making selected comparisons among groups of recipients. Two features of the sampling design for the survey of UI recipients will result in losses of precision, relative to what could be achieved based on a nationwide simple random sample (SRS) of recipients. First, the sample will be clustered into a subset of UI jurisdictions. Second, the sample will be non-proportionally distributed across BYB date ranges and MNW strata due to the sampling objectives described in Section B.1, which will result in variation in the sampling weights used to construct survey estimates. These losses of precision, relative to a nationwide SRS, are commonly referred to as “design effects.”

The design effects from clustering and unequal sampling weights are each described below, followed by a discussion of the implications for the precision of descriptive statistics and subgroup comparisons. The results of the analysis of statistical power presented there suggest that the comparisons of UI-only recipients to extended benefits recipients will reliably reveal fairly small differences. More targeted comparisons of subgroups defined by the BYB calendar quarter will likely be able to statistically detect large, but not modest, differences between groups.

Design Effects from Clustering. A two-stage clustered sample design will yield less precise estimates than an SRS covering the full population of UI recipients. This loss of precision occurs because individual outcomes tend to be more strongly correlated within UI jurisdictions than across jurisdictions. Adding an individual to one of the sampled jurisdictions yields a smaller amount of new information than if an individual from an entirely different jurisdiction were brought into the

survey. Thus, the same amount of information provided by a clustered design could be obtained by sampling fewer individuals in more jurisdictions.

The key factor that determines the extent of the design effect from clustering is the intraclass correlation coefficient (ICC), which measures the proportion of the variability of individual outcomes that can be explained by jurisdiction-specific factors. Corson et al. (1999) present design effects from clustering for a range of characteristics and outcomes of recipients of UI and of EUC. Calculations based on these data suggest an ICC for UC duration of approximately 0.04, ICCs for demographic characteristics ranging from less than 0.01 (age) to 0.07 (race), and ICCs for unemployment duration and reemployment earnings of less than 0.01. Although it may seem negligible that less than 7 percent of variability in individual characteristics and outcomes is attributable to jurisdiction-specific factors, these numbers can actually result in substantial design effects because the sample of UI jurisdictions is much smaller than the sample of individuals.

Design Effects from Unequal Weighting. As explained in Section B.1, the sample of recipients will be allocated evenly, rather than proportionately, across BYB date ranges in the second-stage of the sampling process. Although this will maximize the precision of comparisons between BYB date ranges, it will reduce the precision of overall descriptive statistics that pool information across all BYB dates. This loss of descriptive precision occurs because an even allocation implies that some date ranges are oversampled while others are undersampled. Unequal weights must be applied to obtain representative estimates, thereby increasing the variance of pooled estimates. Intuitively, this design effect from unequal weighting can be thought of as occurring because the extra individuals in an oversampled BYB date range are providing less distinctive information than if additional individuals were instead selected from an undersampled date range.

A similar design effect from unequal weighting results from survey nonresponse because the propensity of nonresponse may vary according to the characteristics of UI recipients. Some types of individuals might be overrepresented in the final survey sample, while others may be underrepresented. As described in Section B.3, the initial weights derived from the sampling will be adjusted accordingly. The extent of the adjustment will also vary according to UI recipient characteristics, resulting in an increase in the variance of the survey estimates.

Consequences of Design Effects for Descriptive Statistics. To summarize the implications of the survey design for the precision of descriptive statistics (for example, means and proportions), Table B.4 includes information on a *combined design effect* for various study populations. This combined effect is calculated as the product of the design effect from clustering and the expected design effect from unequal weighting and may be interpreted as the ratio of the variability of estimates based on the clustered, explicitly stratified design to the variability that would be obtained in an SRS drawn proportionately from the full population of UI recipients. The table includes combined design effects evaluated using a range of plausible ICCs for: (1) the full survey sample; (2) a 50 percent subgroup, which might be thought of as representing one of the study populations (UI-only recipients and extended benefits recipients); and (3) a 25 percent subgroup, which may be thought of as representing the number of UI recipients in each six-month BYB date range.

Considering a 50 percent subgroup, such as what might be used for comparisons, drawn using the clustered, stratified design, the estimated mean for a demographic characteristic with an ICC of 0.01 will have a variance that is about 99 percent larger than what could be obtained with an SRS of the full population of recipients. For an outcome such as UI duration (with an ICC of 0.04), the variance of the survey estimate is expected to be 4.2 times as large as the variance that would be obtained from an SRS. For a variable such as race, for which 7 percent of the variation might be

explained by jurisdiction-specific factors, the clustered, stratified design results in a variance that is over 6.4 times as large as the variance from an SRS.

Another way to describe design effects is in terms of the *effective sample size*. This represents the number of recipients drawn at random from the full population that would be expected to yield the same precision as the actual sample size from the two-stage clustered and stratified sample design. Thus, if the ICC is 0.01, a sample of 604 recipients chosen using an SRS would result in approximately the same precision that can be achieved using 1,200 such individuals in this study's two-stage design. Likewise, with ICCs of 0.04 and 0.07, an SRS would need only 286 and 187 recipients, respectively, to achieve the same level of precision as what is obtained in the clustered, stratified random sample based on 1,200 recipients. Decreasing the number of individuals included in the analysis, such as when considering a 25 percent subgroup, will result in further decreases in precision, as can be seen when comparing the effective sizes across groups in Table B.4.

Table B.4. Design Effects and Effective Sample Sizes for the Two- Stage UI Recipient Sample

Sample	Actual Sample Size	Combined Design Effect	Effective Sample Size
ICC = 0.01			
Full sample	2,400	2.74	877
50 percent subgroup	1,200	1.99	604
25 percent subgroup	600	1.61	372
ICC = 0.04			
Full sample	2,400	7.20	333
50 percent subgroup	1,200	4.20	286
25 percent subgroup	600	2.70	222
ICC = 0.07			
Full sample	2,400	11.66	206
50 percent subgroup	1,200	6.41	187
25 percent subgroup	600	3.79	158

Notes: The combined design effect represents the product of the design effect from clustering and design effects from unequal weighting. The effective sample size is calculated by dividing the actual sample size by the design effect.

ICC = intraclass correlation coefficient.

Minimum Detectible Subgroup Differences. One of the primary purposes of collecting survey data for this study is to enable comparisons between groups of recipients based on the availability and actual utilization of additional weeks of benefits from the EUC08 and EB programs. To assess the degree of precision when making such comparisons, Table B.5 displays illustrative minimum detectible differences (MDDs) for contrasts within the following three sets of subgroups:

- **Contrast 1:** 1,200 recipients in each group spread evenly across all BYB date ranges, which may be thought of as representing a comparison between the UI-only recipients and extended benefits recipients
- **Contrast 2:** 600 recipients in each group spread evenly across all BYB date ranges, which may be thought of representing a comparison between male and female extended benefits recipients

- **Contrast 3:** 300 recipients with BYB dates in one quarter and 300 recipients in the next quarter. Using the last quarter of 2007 and the first quarter of 2008, this contrast may be thought of as representing the contrast between (1) recipients who experienced a gap after UI exhaustion and the availability of extended benefits, due to when the EUC08 program was established in June 2008, and (2) recipients who were able to progress smoothly from UI benefits onto EUC08 benefits. Alternatively, using the third and fourth quarters in 2008, the contrast could represent a comparison between (1) recipients who experienced a gap between EUC08 Tier 2 exhaustion and the availability of Tier 3, due to when Tier 3 was established in November 2009, and (2) recipients who were able to progress smoothly from EUC08 Tier 2 onto Tier 3.

The MDDs have been calculated using standard assumptions about statistical power (80 percent) and the significance level of the test that would be applied (5 percent, two-tailed). The table also focuses on two values of the ICC—0.01 and 0.04—which might be indicative of the extent of clustering in reemployment outcomes and in UI duration, respectively. Finally, the table presents MDDs based on two values for the degree of correlation between outcomes across subgroups within UI jurisdictions: 0.5, which represents a conservative lower bound, and 0.8, which represents a moderate to strong cross-group similarity.

For continuous characteristics, a minimum detectable *standardized difference* for each outcome variable is calculated by dividing the MDD by the standard deviation. This yields a common metric of standard deviation units for expressing differences among groups across all characteristics. A standardized difference of 0.25 is typically regarded as large (see, for example, Institute of Education Sciences 2008). Based on data from Corson et al. (1999), this would translate into a between-group difference in unemployment duration of approximately three months. Many evaluations seek to identify more modest standardized differences on the order of 0.10 to 0.15, which would amount to a difference in unemployment duration of 1.2 to 1.8 months.

As shown in Table B.5, the sample will allow fairly small standardized differences of 0.13-0.16 to be detected for Contrast 1. When considering a binary attribute that is evenly split across the population, for example gender, the MDDs suggest that a statistically significant difference of 6.6 to 8.0 percentage points in the prevalence across groups could be detected, depending on the ICC and on the degree of cross-group within-cluster correlations. For attributes that are relatively uncommon, for example ones that are present for 10 percent of the population, the survey will allow smaller intergroup differences of 3.9 to 4.8 percentage points to be detected for Contrast 1.

When considering Contrast 3, the survey will generally allow large, but not modest, differences between subgroups to be detected, and the results are fairly similar for all values of the correlation parameters. Based on Table B.5, a standardized difference of approximately 0.25-0.26 could be reliably identified, which translates to a difference in UI durations of just over three months. Similarly, a difference of 12.4 to 13.1 percentage points could be detected for binary attribute that is shared by half the population. For an uncommon attribute that has a 10 percent overall prevalence, between-group differences would need to be larger than 7.4-7.9 percentage points to be statistically detected.

Table B.5. Minimum Detectible Subgroup Differences

Comparison	Minimum Detectible Standardized Difference	Minimum Detectible Difference in Percentage Points for a Binary Outcome with an Overall Incidence of:		
		10 percent	25 percent	50 percent
ICC = 0.01; r = 0.5				
Contrast 1	0.136	4.1	5.9	6.8
Contrast 2	0.187	5.6	8.1	9.3
Contrast 3	0.251	7.5	10.9	12.6
ICC = 0.01; r = 0.8				
Contrast 1	0.131	3.9	5.7	6.6
Contrast 2	0.183	5.5	7.9	9.1
Contrast 3	0.248	7.4	10.8	12.4
ICC = 0.04; r = 0.5				
Contrast 1	0.159	4.8	6.9	8.0
Contrast 2	0.203	6.1	8.8	10.1
Contrast 3	0.262	7.9	11.3	13.1
ICC = 0.04; r = 0.8				
Contrast 1	0.140	4.2	6.1	7.0
Contrast 2	0.188	5.6	8.1	9.4
Contrast 3	0.250	7.5	10.8	12.5

Notes: Minimum detectable standardized differences were calculated based on effective sample sizes that take into account the expected design effects from unequal weighting and that apply equations (1) and (10) from Schochet (2005). The latter equation has been modified to allow for unequal effective sample sizes. In addition, all calculations are based on the following assumptions: 80 percent level of power; a two-tailed test at a 5 percent significance level; 9 certainty jurisdictions that contain 42 percent of the study population; 11 noncertainty jurisdictions that contain 58 percent of the study population.

ICC = intraclass correlation coefficient; r = between-group, within-cluster correlation in outcomes.

Other Factors Affecting Precision. In the first stage of the sampling process jurisdictions from the low- and medium-MNW strata are to be oversampled, while jurisdictions with high values of the MNW variable are to be undersampled. To the extent possible, the second-stage allocation of the sample of recipients will compensate for such deviations from proportional sampling in the first stage by allocating fewer claimants to the oversampled states and more claimants in the undersampled states. Any remaining design effect arising from the need to apply unequal weighting across strata would reduce the precision of overall (pooled) descriptive statistics and, to a lesser extent comparisons across BYB date ranges. In addition, if cluster-robust or cluster-bootstrap methods are used to conduct statistical inference, the improvement in Type I error is expected to reduce the statistical power of the test for any pre-specified effect size. Thus, applying such methods will result in less precision, and therefore larger MDDs, than what are presented in Table B.5. Offsetting this might be a gain in precision achieved by implementing an adjustment that accounts for the degree of variability in the first-stage sampling distribution. Finally, it is not clear how covariates included in the estimating equations (6)-(8) will affect precision of the estimates. Precision will go down in cases where the other control variables are more strongly correlated with the explanatory policy variable than they are with the outcome variable, and precision will go up if the opposite is true.

3. Methods to Maximize Response Rates and Data Reliability

The methods to maximize response and data reliability are discussed for each data collection effort that is part of this request for clearance in the subsections below.

a. Response Rates for the UI Recipient Survey

This study has two levels of potential nonresponse: the UI jurisdiction and the selected individual UI recipients in a state. Established procedures to maximize response rates at both levels will be followed, as described below. Strategies to address potential nonresponse bias are discussed in the next subsection.

Maximizing Jurisdiction-Level Response Rates. While the study aims to achieve 100 percent cooperation among the UI jurisdictions selected for inclusion in the sample, some states may refuse to provide the claims data needed to locate UI recipients. The study will maximize jurisdictions' participation by adopting practices employed in previous successful recruitment efforts. In the recent Impact Evaluation of the Trade Adjustment Assistance Program (TAA study), the evaluation team requested that states deliver large, multipart UI administrative data files in 2010, after the end of the recession. UI claims and wage data were successfully obtained from all 26 states that were contacted for the TAA study. This study of the UC provisions of ARRA will use similar state recruitment methods, including coordinating recruitment efforts between DOL and the contractor, formulating as simple a data request as possible, and offering logistical support and cost-recovery payments to UI jurisdictions.

Maximizing Individual-Level Response Rates. The strategy for maximizing response to the UI recipient survey will be based on the approaches described below, which have been successfully used in many other studies. The methods employed will address all types of individual nonresponse, including failure to locate the sample member or his or her refusal to participate in the survey.

Contact with sample members. The contractor will send an advance letter on DOL letterhead to sample members before attempting to contact them by phone. This letter will (1) introduce the study and its purpose, (2) highlight DOL as the study sponsor, (3) explain the voluntary and private nature of participation, (4) extend the incentive offer, (5) provide web survey log-in information, and (6) give a toll-free number for telephone calls. The envelope will be printed with the DOL logo to capture the sample members' attention and to communicate the legitimacy of the study. The research contractor's return address will be used to facilitate the processing of returned mail and locating procedures. An information sheet providing answers to questions that sample members may have about the study will be included with the advance mailing. It also will include a phone number and a DOL website address that sample members can use to learn more about the study. The advance letter will be followed up with timed reminders offering the option to complete the survey via the telephone or the web. Copies of the advance letter, FAQs, and reminders (postcards and letter) that will be sent to sample members are included as Appendix F.

Before the mailing of these materials, interviewing staff, such as interviewers, project supervisors, monitors, and locators, at Mathematica's Survey Operations Center (SOC) will be thoroughly trained on how to address respondents' questions about the study and questionnaire. In addition to the sheet of answers to questions that will accompany the advance mailing, a more extensive list of frequently asked questions and answers (FAQs) will be developed for the interviewers' use. These FAQs will be included in the operational procedures manual for the computer-assisted telephone interviewing (CATT)-administered questionnaire, and integrated into

the CATI instrument. Interviewers will be able to access the FAQs at any time during the interviewer-administered survey. Other FAQs will be available online for the self-administered web survey and web survey respondents will have access to them throughout the survey.

Locating sample members. A key component to obtaining a high response rate is locating sample members. The process of locating UI recipients selected for the study will begin before sending out the first mailing. This locating process will involve the use of an independent vendor that will check the full sample against current address databases. This first step is critical given that (1) the contact information for some sample members may be from as far back as late 2007 and (2) some sample members may have moved. Extensive tracking and locating procedures that have proven successful in other Mathematica studies will be used for sample members whose mail is returned as undeliverable. These include using other independent databases, checking with neighbors and family members, and searching social networking sites. When talking with contacts, the specific purpose of the call will not be disclosed, but it will be stated that the effort to reach the sample member is for an important study being sponsored by the government.

Gaining and maintaining cooperation. A key component to achieving high response rates is gaining cooperation after locating respondents. Mathematica's interviewers are highly trained in establishing rapport with gatekeepers, gaining cooperation, and avoiding refusals. Sample members who are difficult to contact and who have not yet completed the survey on the web will be sent a reminder postcard one week after the advance letter and a follow-up postcard two weeks later. A reminder letter will be sent mid-way through the data collection period and again three to four weeks before the end of data collection to remaining nonrespondents. To those sample members who refuse to participate, a targeted refusal conversion letter that will address their specific concerns will be mailed first. Next, expert refusal conversion interviewers will make follow-up calls to try to gain the sample members' cooperation.

Multi-language survey administration. During phone contact, interviewers will identify Spanish-speaking respondents and connect or schedule them to speak with a bilingual interviewer. When necessary, translators for languages other than Spanish will be used; Mathematica employs staff who speak a wide range of languages and have experience conducting interviews in a number of languages.

Incentives for survey participants. Offering an incentive for the UI recipient survey is essential to generate the desired response rates and reduce overall survey costs without affecting data quality. There is substantial evidence on the benefits of offering incentives. According to Singer et al. (2000), incentives can help achieve high response rates by increasing the sample members' propensity to respond; by doing so, incentive payments have been found to contain evaluation costs by significantly reducing the number of calls required to resolve a case. Studies offering incentives show decreased refusal rates and increased contact and cooperation rates. Incentives also increase the likelihood of participation from subgroups with a lower propensity to cooperate with the survey request. This is an important component of ensuring the representativeness of the survey respondents and the quality of the data being collected. For example, Jäckle and Lynn (2007) find that incentives increase the participation of sample members more likely to be unemployed. There is also evidence that incentives bolster participation among those with lower interest in the survey topic (Schwartz et al. 2006; Jäckle and Lynn 2007; Kay 2001), resulting in data that are more nearly complete. Furthermore, paying incentives does not impair the quality of the data obtained (such as item nonresponse or the distribution of responses) from groups who would otherwise be underrepresented in the survey (Singer et al. 2000).

An incentive will be offered to all survey respondents, using a two-tiered incentive offer to encourage the selection of the less expensive web option for survey administration—\$50 for completion on the web and \$40 for completion using CATI. Based on the pervasive use of the web by a cross-section of the general population, it is anticipated that a substantial number of sample members will choose the web, since many of them are likely to be more comfortable with this self-paced, self-administered approach. Also, the higher incentive offer for web completion will encourage many to use that option. In the National Survey of Recent College Graduates, conducted by Mathematica for the National Science Foundation, approximately 20 percent more survey completions were obtained when sample members were offered a \$30 incentive instead of \$20. The web survey will be available as soon as invitations are mailed to sample members. It is estimated that 40 percent of the completed surveys will come from the web.

To leverage fully the benefits of offering incentives in the UCP evaluation, the advance letter to the UI study participants will mention the incentive. Interviewers will also mention the proposed incentive when they establish contact with the participants and attempt to gain their cooperation.

Survey length. The UI recipient survey questionnaire is designed to be easy to complete. The questions are written in clear and straightforward language. The average time required for the respondent to complete the survey, either on the web or by telephone, is estimated at 30 minutes.

Interviewer training. Mathematica has a cadre of survey operations staff who are experienced working on previous studies conducted for DOL as interviewers, supervisors, and monitors. These staff are familiar with similar questionnaire content and are sensitive to the difficulties faced by jobseekers and unemployed individuals. To the extent possible, Mathematica will assign these experienced staff to the UCP evaluation. All survey operations staff assigned to the study will participate in general training (if not already trained) as well as extensive project-specific training. Interviewers will not work on the study until they have been certified as prepared. The project-specific training will include role playing with scenarios and other techniques to ensure that interviewers are ready to respond effectively to sample members' questions. Responses to frequently asked questions will be reviewed, as will each questionnaire item. Interviewers will participate in supervised paired-practice sessions before they are certified as ready to interview for the project. Training sessions will stress the importance of being sensitive to respondent's situations while remaining impartial. They will also focus on developing skills for securing respondents' cooperation and averting and converting refusals.

Targeted response rate. Employing these procedures, an 80 percent response to the UI recipient survey is targeted. When the survey is completed, an analysis that compares respondents to nonrespondents will be conducted to assess whether the survey sample is representative of the target population of UI recipients. This analysis will be done using UI claims and wage record data, which will be available for all sample members. These data will include demographic variables (sex, age, race/ethnicity), earnings measures (base period earnings and quarterly earnings from the UI wage records), and UI claim data (WBA, maximum benefit amount, weeks collected, and dollars collected). If it appears that the survey respondent sample is not representative, sample weights will be adjusted for nonresponse using propensity scoring methods.

b. Nonresponse Bias Analyses for the UI Recipient Survey

A bias may arise in study results if participating jurisdictions and individuals differ from the target population as a whole. The nonresponse bias analysis will provide some indication of whether a possible nonresponse bias exists and the data items and populations for which survey estimates

might have a greater potential for bias. However, because survey data will not be available for nonrespondents, the analysis can never determine conclusively if bias does or does not exist in the survey estimates.

Nonresponse Bias Analysis at the Jurisdiction Level. Jurisdiction-level nonresponse results in the exclusion of a relatively large number of people, and the reason for the refusal of the jurisdiction to provide data may be correlated with the outcomes of interest for this evaluation. To assess the possibility of bias arising from jurisdiction-level nonresponse, both qualitative and a quantitative analyses will be conducted.

The qualitative analysis will concentrate on the reasons for refusal given by UI jurisdictions that choose not to cooperate with the data request. Of particular concern is whether economic conditions or policies that could affect the outcomes of interest for this evaluation play a role in a refusal to provide data because this may indicate a potential for bias. The results of the qualitative analysis could be consistent with the expectation that UI jurisdictions experiencing more strain on their UC system due to the recession are less likely to cooperate with a data request. In that case, the first-stage stratification system described in Section B.1 would be expected to mitigate the potential bias arising from differences across jurisdictions in the increase in UI claims stemming from recessionary strains. Depending on the results of the quantitative analysis described below, this could increase the confidence with which the study team might be able to make robust inference about the national population of UI claimants using the sample of jurisdictions selected for this study. Alternatively, if UI jurisdictions identify other economic factors or policies as being more salient in a refusal decision, these could be included as variables in the quantitative analysis.

The quantitative analysis will have two components:

1. The study team will examine the extent to which the attributes of noncooperating jurisdictions differ systematically from the attributes of cooperating jurisdictions. This analysis will examine jurisdiction-level data available from DOL on the number of UI claims, number of first payments, and total benefits paid out on a monthly basis. The analysis will also consider differences across jurisdictions in the policies identified in the qualitative analysis.
2. Estimates from the Current Population Survey (CPS) can be used to compare the distribution of characteristics of the UI recipient population in responding jurisdictions to the full set of selected jurisdictions using the individual-level analysis methods described in the next subsection.²⁵ Some of the characteristics available from the CPS include age, race/ethnicity, gender, occupation, and industry.

Each of these analyses can provide suggestive evidence on the extent to which jurisdiction-level response varies according to characteristics that are likely to be significant predictors of the outcomes of interest for this study. As such, the results from the nonresponse bias analysis could affect the study's conclusions.

²⁵ Measures derived from the CPS will be calculated using the sampling weights provided in that survey.

Substantive differences between cooperating and noncooperating jurisdictions, and/or strong associations between outcomes and nonresponse-relevant economic factors within the cooperating jurisdictions would indicate nonresponse that would be considered “informative,” relative to the potential outcomes of the sample members. Informative nonresponse would suggest a form of selection bias at the jurisdiction level, in which case it would not be reasonable to calculate fully nationally representative estimates using the survey sample. We will assess multiple ways to analyze these data. In one approach, the study team could seek to conduct design-based inference about a population of UI jurisdictions that the sample most closely resembles (that is, a population of UI jurisdiction with a similar distribution of the characteristics found to be significant in the analyses described above). In this case, inference could be based only on the main sample or on the entire augmented sample (including jurisdictions from the main and reserve samples), depending on the results of the qualitative analysis. Estimates based on this approach would be presented with appropriate cautions regarding the extent to which the findings can actually be generalized to such a population. Second, the study team could simply treat the entire augmented sample of cooperating jurisdictions as a convenience sample. In this case, statistical inference would be valid within the sample only, and the presentation of the findings would make it clear that estimates based on such an analysis do not generalize to any clear population.

If the quantitative analyses of jurisdiction-level nonresponse do not yield significant results (i.e., “uninformative” nonresponse), this suggests that selective nonresponse is less likely to introduce bias in the study’s findings. In this case, the study team would use the main or augmented sample (depending on the results of the qualitative analysis) to calculate national estimates. However, the study would explicitly acknowledge that (1) estimates could still be biased based on factors not accounted for in the quantitative nonresponse analysis and (2) the relatively small sample size of UI jurisdictions could limit the power of the quantitative analysis to reveal statistical differences. The findings of the study would include appropriate caveats for readers.

Nonresponse Bias Analysis at the Individual Level. As with almost any survey, some nonresponse among the UI recipients selected for the study is inevitable. Some sample members will not be located and others will not be able or willing to respond to the survey. The nonresponse bias analysis will use various data items in the administrative data files, including demographic information, employment status and quarterly earnings. The nonresponse bias analysis will consist of the following steps:

1. Compute response rates for key subgroups.
2. Compare the distributions of respondent and nonrespondent characteristics using initial sampling weights.
3. Identify the characteristics that best predict nonresponse and use this information to generate nonresponse weight adjustments.
4. Post-stratify survey estimates of the size of the study population to match national totals.
5. Compare the distribution of characteristics of respondents using the fully response-adjusted analysis weights to the distribution of characteristics of the full sample using the unadjusted sampling weights.

These bias analyses will build on the individual-level nonresponse analysis used to adjust the survey sampling weights to compensate for this nonresponse (see Section B.1). The analyses will be

conducted within and across UI jurisdictions to assess whether the potential for nonresponse bias differs among jurisdictions. Each of these steps is discussed below in greater detail.

Compute response rates for subgroups. The response rate for the subgroups will be computed using the American Association for Public Opinion Research definition of the response rate: the weighted number of completed interviews with eligible participants divided by the estimated number of eligible individuals (AAPOR 2011). Overall response rates will be computed for the full sample and by jurisdiction. Response rates will then be computed for subgroups defined by characteristics available in the UI claims data to examine if these rates differ systematically from the overall response rate.

Compare the characteristics of respondents and nonrespondents. Next, the characteristics of respondents and nonrespondents will be calculated according to characteristics available in the UI claims data. The statistical significance of the difference between the respondent and nonrespondent subgroups will be assessed using *t*-tests. This type of analysis can be useful in identifying patterns of differences in observable characteristics that might suggest nonresponse bias, it can be affected by small sample sizes and generally has low power to detect substantive differences. The large number of statistical tests conducted can also result in high rates of Type I error.

Identify the best explanatory factors of nonresponse and generate nonresponse weight adjustments. As described in Section B.1, logistic regression modeling is commonly used to develop adjustment factors for nonresponse. This approach is also known as response propensity modeling and can be viewed as an extension of the classical weighting-class nonresponse adjustment procedure that makes it possible to include more factors (that is, binary, categorical, and continuous factors) in nonresponse adjustments. A CHAID analysis will be used to assist in identifying potentially significant interactions among the subgroups or factors available for all individuals. The final response propensity model will be using variables developed from the interaction terms identified in the CHAID analyses. Based on the final model, the inverse of the predicted propensity to respond will be used as an adjustment factor to the initial sampling weights.

Computing nonresponse adjustment factors will contribute substantially to the nonresponse bias analysis by identifying the main effects and interaction among main effects that are statistically associated with nonresponse. This information will be used in the bias analysis to form levels of categorical variables for computing response rates and point estimates using both the original sampling weights and the nonresponse adjusted sampling weights.

Post-stratify survey estimates to match available national totals. Post-stratification is a procedure whereby the response-adjusted weights are further adjusted so that survey estimates of the size of the study population are aligned to known totals external to the survey. This process offers face-validity for reporting population counts and has some statistical benefits. In this study, survey estimates of the number of UI recipients with first payments in each BYB date range will be post-stratified to the national counts available from ETA.

Compare the fully-adjusted weighted distribution of respondent characteristics to the distribution for the full sample using initial weights. In this last step, the distribution of respondent baseline characteristics will be compared to the distribution for the full study population and for key subgroups. This analysis can highlight measures where the potential for nonresponse bias is greatest and where greater caution should be exercised in the interpretation of the observed findings.

c. Reliability of Data Collection for the UI Recipient Survey

The UI recipient survey includes questions that have been widely used and tested in the field by other recent studies such as the Trade Adjustment Assistance Study Follow-Up Survey (OMB number 1205-0460) and the Individual Training Account 2 (ITA2) Follow-up Questionnaire (OMB 1205-0441). Other surveys referenced were the Temporary Extended Unemployment Compensation questionnaire, the UI Exhaustee questionnaire, and the Emergency Unemployment Compensation questionnaire. During development, the UI recipient survey questionnaire was reviewed by staff at DOL, Mathematica project staff, and members of the project's Technical Working Group (TWG). The survey has also been pretested with UI recipients. [NOTE: This section will be updated after completion of pretests.]

In addition, to better understand the reliability of the data reported, differences in answers across modes (web or CATI) will be carefully reviewed. In the web survey, an answer to key questions will be required before the respondent can proceed; programming the instrument this way will improve the completeness of data and, hence, the response rate.

Further, because it is expected that some sample members will have multiple UI benefit years in the period of interest, the study researchers will establish the UI claim date of interest at the beginning of the survey. Other recall aids such as dates of employment subsequent to job loss and dates of enrollment in school or training program, will be recorded and retained by the CATI and web programs and used at appropriate questions. Probes, verifications, and consistency checks will be programmed into both the CATI and web versions of the survey, further ensuring the reliability of the data collected. Except for language necessary to accommodate self-administration versus being asked by an interviewer, the content of both survey versions will be identical.

Finally, interviewing supervisors will monitor at least 10 percent of each interviewer's work using silent call-monitoring equipment and video monitors that display the interviewer's screen. Supervisors will evaluate interviewer performance based in part on this monitoring. Supervisors will then discuss these evaluations and coach interviewers to ensure high-quality data collection. Retraining and/or re-assignments will be provided as needed.

d. Response Rates for the Survey of UI Administrators

State UI directors in the 50 states and the District of Columbia will be asked to complete the survey of UI administrators or to have a designee do so. A high response rate (targeted to be 100 percent) will be achieved through strategies that facilitate easy completion of the survey. One such strategy is that the survey is designed as a self-administered questionnaire that can be completed on the hard copy or by computer using a write-enabled pdf-formatted questionnaire; this will allow the administrators to complete the survey at a time that is convenient for them. More generally, Mathematica will mail a letter of invitation and a survey booklet to the 51 UI administrators asking for their participation. Also, an electronic version of the questionnaire will be emailed to administrators for whom email addresses are available or upon their request. To ensure maximum flexibility for the respondent, UI administrators can email or fax the completed survey back or return it via regular mail using a pre-paid business reply envelope that will be included with the initial mailing packet.

In addition, the survey of UI administrators begins with a pre-filled, state-specific fact sheet pertaining to the state's adoption of the UC-related ARRA provisions. Administrators will be asked to confirm or correct the pre-filled information. The remainder of the survey is identical for all

administrators, but it contains appropriate skip logic and instructions so that non-applicable questions can be disregarded. Taken together, use of the pre-filled information and the skip logic in the questions will ensure that the survey can be completed without undue burden on the respondents. The survey is expected to take an average of about 30 minutes to complete.

The study team will contact state administrators who do not respond within two weeks to encourage them to do so, finding ways to make participation as easy as possible for UI administrators. TWG members or members of professional associations may be contacted to aid in the effort to secure survey cooperation.

e. Reliability of Data Collection for the Survey of UI Administrators

Several strategies to ensure the reliability of the data collected through the survey of UI administrators will be used. First, a state-specific fact sheet that is pre-populated with publicly available information about the adoption of UC provisions (discussed above) will be provided to administrators, with a request that they confirm or correct the information. In addition, administrators will be encouraged to collaborate with colleagues and/or to delegate completion of specific questionnaire items as needed.

When the completed surveys are returned, they will be reviewed by project staff, who will follow up as appropriate with the main respondent (or his or her designee) for clarification or to request responses to any incomplete items.

f. Response Rates for the Site Visit Data Collection Effort

The plan to collect study data during site visits will ensure that response rates are high and that the data are reliable. After receiving DOL approval of the 20 states selected for the study, a letter will be sent to each state's UI director introducing the study, informing the director of the interest in visiting the state, and indicating that a researcher will call to schedule an initial phone call.²⁶ During this initial phone call, the study researcher will explain the purpose of the study so the UI director will be aware of what is expected upon agreeing to participate. The study researcher also will obtain information on which staff within the UI office would be best able to respond to the various protocol modules; solicit suggestions about other stakeholders, such as legislators, lobbyists, and advisory council members to contact for interviews; and identify possible visit dates. Before the initial phone call to the UI director, the researcher assigned to work with each state will review publicly available background materials and responses to the survey of UI administrators to discern which optional provisions were adopted and the political and economic context of the state. This information will enable the researcher to verify any information that is unclear and determine which respondent categories will be targeted during the site visit. In a follow-up email, the site visitor will thank the administrator for agreeing to participate in the research and will also summarize the purpose of the visit and relay a tentative visit schedule based on information gathered during the discussion.

²⁶ Because the administrative data collection and survey of UI administrators will have occurred before the initial phone calls for the site visits, the contractor will coordinate communications with UI directors to inform them of the various study components and explain that the state might be contacted for the study's in-person data collection about decisionmaking and implementation.

Site visitors will begin working with state staff well in advance of each visit to ensure that the timing of the visit is convenient. The site visits will take place over a period of several months, which also will allow flexibility in timing. Because the visits will involve several interviews and activities each day, there will be flexibility in the scheduling of specific interviews and activities to accommodate the particular needs of respondents.

Two weeks before the site visit, the data systems survey will be mailed to the benefits chief (or other appropriate staff member, as identified by the UI director) for completion. The questionnaire is composed primarily of closed-ended questions and will take an average of 45 minutes to complete. The completed questionnaires will be reviewed by the study team before the site visits, and the site visitors will ask clarifying and follow-up questions during the visit.

Each site visit will include both one-on-one and small group interviews, as appropriate and following the guidance of the UI director. For instance, in some states, a one-on-one interview with the UI director might be conducted, while in other states, the same topics might be covered with the UI director and/or top deputies. Should scheduling conflicts prevent a meeting with all respondents while on site, follow-up phone calls will be conducted accordingly. Similarly, should follow-up questions arise after a visit, researchers will call or email respondents for clarification.

g. Reliability of Data Collection for the Site Visits

Four well-proven strategies will be used to ensure the reliability of the data. First, a pilot site visit will be conducted by two experienced site visitors. During this visit, the site visitors will assess the flow and pacing of the discussion that is guided by the questions in the site visit protocol to ensure that it is feasible during a visit to collect comprehensive information that is in accord with the study's goals. As needed, revisions to the protocol will be made to facilitate the data collection effort. Second, all site visitors, most of whom already have extensive experience with this data collection method, will be thoroughly trained in the issues of importance to this particular study. This training will include techniques to probe for additional details to help interpret responses to interview questions and to assure all interview respondents of the privacy of their responses to questions. Third, when appropriate, the protocols will use standardized checklists to further ensure that the information is collected systematically. Finally, each site visit report will be read by a senior member of the evaluation team to ensure that the relevant data are collected and recorded.

4. Tests of Procedures or Methods

All procedures, instruments, and protocols to be used in the conduct of the UCP evaluation will be tested to assess the data collection processes can be evaluated, to evaluate the clarity of the questions to be asked, to identify possible modifications to either question wording or question order that could improve the quality of the data, and to estimate respondent burden. The tests for each of the three data collection efforts that are part of this request for clearance will be conducted prior to the full roll-out of the data collection effort.

UI recipient survey. The UI recipient questionnaire will be thoroughly pretested with nine UI recipients. Following each pretest, project staff will debrief with the participant using a standard debriefing protocol to determine if any words or questions were difficult to understand and answer. If problems are found with the content or timing of the questionnaire, adjustments will be made. [NOTE: This section will be updated after the pretest.] Pretests will be conducted using hard copy versions of the instrument. However, before fielding the survey, rigorous usability tests of both the CATI and web versions will be conducted. Project and survey operations staff will log into CATI

and web test sites to implement different scenarios designed to ensure that all skip logic, fills, layout, response formats, and overall survey navigation pass stringent requirements.

Survey of UI administrators. Since the survey will be administered to all 50 states and the District of Columbia, the options for conducting a pretest of the survey are limited. Asking a state to complete the survey as a pretest might affect its responses when the survey is deployed. Thus, the team will ask one member of the study’s TWG who is a former UI state administrator to complete the survey as he would in that capacity and to provide us with his feedback on the content and length of the survey. The team will also ask other members of the TWG for their suggestions of other individuals who currently are not state administrators but who could provide us with appropriate feedback on the survey.

Site visits. To ensure that the site visit protocols are used effectively as field guides and that they yield comprehensive and comparable data across the 20 states, senior research team members will conduct a pilot site visit before the full round of site visits. The purposes of the pilot test are to ensure that the field protocol, which will guide field researchers as they collect data on site, include appropriate probes that assist site visitors in delving deeply into topics of interest and that the protocols do not omit relevant topics of inquiry. Furthermore, use of the protocols during a pilot site visit can enable the research staff leading this task to assess that the site visit agenda that the research team develops—including how data collection activities should generally be structured during each site visit—is practical given the amount of data to be collected and the amount of time allotted for each data collection activity. Adjustments to the site visit guides will be made as necessary.

5. Individuals Consulted on Statistical Methods

To ensure that the best decisions were made regarding the statistical aspects of the design, experts from outside the agency were consulted, and their input has helped to shape the sampling design. These experts included project staff from Mathematica and the Urban Institute, as well as members of the project’s TWG. The experts consulted are listed below, along with telephone contact information. Only evaluation staff from Mathematica and the Urban Institute will collect and analyze the information.

Mathematica

Dr. Karen Needels, Project Director	(541) 753-0201
Dr. Walter Nicholson, Co-Principal Investigator	(413) 542-2191
Ms. Linda Rosenberg, Task Leader-States’ Decision-Making Analysis and Implementation Study	(609) 936-2762
Dr. Frank Potter, Senior Fellow	(609) 936-2799
Dr. Eric Grau, Senior Statistician	(609) 945-3330
Dr. Heinrich Hock, Research Economist	(202) 250-3557
Dr. Annalisa Mastro, Senior Researcher	(609) 275-2390

The Urban Institute

Dr. Wayne Vroman, Co-Principal Investigator	(202) 261-5573
---	----------------

Members of the Technical Working Group (TWG)

Dr. Rich Hobbie, National Association of State Workforce Agencies	(202) 434-8020
Dr. Till von Wachter, Russell Sage Foundation and Columbia University	(212) 355-3406
Dr. Stephen Woodbury, Michigan State University and W. E. Upjohn Institute for Employment Research	(269) 385-0408

REFERENCES

- American Association for Public Opinion Research (AAPOR). 2011. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Seventh edition. Deerfield, IL: AAPOR.
- ATLAS.ti. *Qualitative Software Data Analysis*. Berlin, Germany: ATLAS.ti GmbH, 2011. Available at [http://www.atlasti.com/]. Accessed May 10, 2011.
- Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. “How Much Should We Trust Differences-in-Differences Estimates?” *Quarterly Journal of Economics*, vol. 119, no. 1, February 2004, pp. 248–275.
- Binder, D. A. “On the Variances of Asymptotically Normal Estimators from Complex Surveys.” *International Statistical Review*, vol. 51, 1983, pp. 279–292.
- Biggs, David, Barry de Ville, and Ed Suen. “A Method of Choosing Multiway Partitions for Classification and Decision Trees.” *Journal of Applied Statistics*, vol. 18, no. 1, 1991, pp. 49–62.
- Cameron, A. Colin, Jonah Gelbach, and Douglas Miller. “Bootstrap-Based Improvements for Inference with Clustered Errors.” *Review of Economics and Statistics*, vol. 90, no. 3, 2008, pp. 414–427.
- Chromy, James R. “Sequential Sample Selection Methods.” *Proceedings of the American Statistical Association, Survey Research Methods Section*, 1979, pp. 401–406.
- Cook, Thomas D., and Vivian C. Wong. “Empirical Tests of the Validity of the Regression Discontinuity Design.” *Annales d’Economie et de Statistique*, 2008.
- Corson, Walter, Jean Grossman, and Walter Nicholson. “An Evaluation of the Federal Supplemental Compensation Program.” *Unemployment Insurance Occasional Paper No. 86-3*. Washington, DC: U.S. Department of Labor, Employment and Training Administration, 1986.
- Corson, Walter, David Horner, Valerie Leach, Charles Metcalf, and Walter Nicholson. “A Study of Recipients of Federal Supplemental Benefits and Special Unemployment Assistance.” Princeton, NJ: Mathematica Policy Research, January 1977.
- Corson, Walter, Karen Needels, and Walter Nicholson. “Emergency Unemployment Compensation: The 1990s Experience Revised Edition.” *Unemployment Insurance Occasional Paper No. 99-4*. Washington, DC: U.S. Department of Labor, Employment and Training Administration, 1999.
- Folsom, Ralph E., Francis J. Potter, and Steven R. Williams. “Notes on a Composite Measure for Self-Weighting Samples in Multiple Domains.” *Proceedings of the American Statistical Association, Survey Research Methods Section*. Alexandria, VA: American Statistical Association, 1987, pp. 792–796.
- Froot, Kenneth A. “Consistent Covariance Matrix Estimation with Cross-Sectional Dependence and Heteroskedasticity in Financial Data.” *Journal of Financial and Quantitative Analysis*, vol. 24, no. 3, 1989, pp. 333–355.
- Heckman, James J. “Sample Selection Bias as a Specification Error.” *Econometrica*, vol. 47, no. 1, 1979, pp. 153–161.

- Institute of Education Sciences, U.S. Department of Education. *What Works Clearinghouse: Procedures and Standards Handbook (Version 2.0)*. 2008. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/wwc_procedures_v2_standards_handbook.pdf on September 8, 2010.
- Jäckle, Annette, and Peter Lynn. “Respondent Incentives in a Multi-Mode Panel Survey: Cumulative Effects on Nonresponse and Bias.” Working paper presented to the Institute for Social and Economic Research, University of Essex, Colchester, United Kingdom, 2007.
- Kalbfleisch, John D., and Ross L. Prentice. *The Statistical Analysis of Failure Time Data*, 2nd edition. Hoboken, NJ: John Wiley and Sons, 2002.
- Katz, Jack, and Arnold Ochs. “Implications of Potential Duration Policies.” In *Report of the National Commission on Unemployment Compensation*, Vol. I. Washington, DC: U.S. Department of Labor., 1980.
- Kass, G. V. “An Exploratory Technique for Investigating Large Quantities of Categorical Data.” *Applied Statistics*, vol. 29, no. 2, 1980, pp. 119–127.
- Kay, Ward R. “The Use of Targeted Incentives to Reluctant Respondents on Response Rates and Data Quality.” Proceedings of the American Association for Public Research. Montreal, Canada: American Association for Public Opinion Research, 2001.
- Kruse, Douglas L. “International Trade and the Labor Market Experience of Displaced Workers.” *Industrial and Labor Relations Review*, vol. 41, no. 3, April 1988, pp. 402–417.
- Magidson, Jay. *SPSS for Windows CHAID Release 6.0*. Belmont MA: Statistical Innovations, Inc., 1993.
- McCall, Brian, and Wei Chi. “Unemployment Insurance, Unemployment Durations and Re-employment Wages.” *Economics Letters*, vol. 99, no. 1, April 2008, pp. 115–118.
- Meyer, Bruce. “Unemployment Insurance and Unemployment Spells.” *Econometrica*, vol. 58, no. 4, July 1990, pp. 757–782.
- Moffitt, Robert. “The Effect of Duration of Unemployment Benefits on Work Incentives: An Analysis of Four Data Sets.” UI Occasional Paper 1985-4. Washington, DC: U.S. Department of Labor, Employment and Training Administration, 1985.
- Needels, Karen, Walter Corson, and Walter Nicholson. “Left Out of the Boom Economy: UI Recipients in the Late 1990s.” ETA Occasional Paper No. 2002-03. Washington, DC: U.S. Department of Labor, Employment and Training Administration, Office of Policy Development, Evaluation and Research, 2002.
- Newton, Floyd C., and Harvey S. Rosen. “Unemployment Insurance, Income Taxation, and Duration of Unemployment: Evidence from Georgia.” *Southern Economic Journal*, vol. 45, no. 3, January 1979, pp. 773–784.
- O’Donnell, Owen, Eddy Van Doorslaer, Adam Wagstaff, and Magnus Lindelow. *Analyzing Health Equity Using Household Survey Data: A Guide to Techniques and Their Implementation*. Washington, DC: World Bank Publications, 2008.

- Potter, Francis J., Vincent G. Iannacchione, William D. Mosher, Robert E. Mason, and Jill D. Kavee. "Sample Design, Sampling Weights, Imputation, and Variance Estimation in the 1995 National Survey of Family Growth." In *Vital and Health Statistics*, series 2, no. 124. Hyattsville, MD: National Center for Health Statistics, 1998.
- Potter, Francis J. "The Effect of Weight Trimming on Nonlinear Survey Estimates." In *Proceedings of the American Statistical Association, Section on Survey Research Methods*. Alexandria, VA: American Statistical Association, 1993, pp. 758-763..
- Potter, Francis J. "A Study of Procedures to Identify and Trim Extreme Sampling Weights." In *Proceedings of the American Statistical Association, Section on Survey Research Methods*. Alexandria, VA: American Statistical Association, 1990, pp. 225-230.
- Rosenbaum, Paul R., and Donald B. Rubin. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika*, vol. 70, no. 1, April 1983, pp. 41–55.
- Research Triangle Institute. *SUDAAN Language Manual, Release 9.0*. Research Triangle Park, NC: Research Triangle Institute, 2004.
- Särndal, Carl-Erik, Bengt Swensson, and Jan Wretman. *Model-Assisted Survey Sampling*. New York: Springer-Verlag, 1992.
- Schochet, Peter Z. "Statistical Power for Random Assignment Evaluations of Education Programs." Report submitted to the U.S. Department of Education, Institute of Education Sciences. Princeton, NJ: Mathematica Policy Research, January 2005.
- Schochet, Peter Z. "Statistical Power for Regression Discontinuity Designs in Education Evaluations." *Journal of Educational and Behavioral Statistics*, vol. 34, no. 2, June 2009, pp. 238–266.
- Schwartz, Lisa K., Lisbeth Goble, and Edward M. English. "Counterbalancing Topic Interest with Cell Quotas and Incentives: Examining Leverage-Salience Theory in the Context of the Poetry in America Survey." *Proceedings of the American Association for Public Research*. Montreal, Canada: American Association for Public Opinion Research, 2006.
- Singer, Eleanor, John Van Hoewyk, and Mary P. Maher. "Experiments with Incentives in Telephone Surveys." *Public Opinion Quarterly*, vol. 64, no. 2, summer 2000, pp. 171–188.
- Wooldridge, Jeffrey. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press, 2002.

MATHEMATICA
Policy Research, Inc.

www.mathematica-mpr.com

Improving public well-being by conducting high-quality, objective research and surveys

Princeton, NJ ■ Ann Arbor, MI ■ Cambridge, MA ■ Chicago, IL ■ Oakland, CA ■ Washington, DC

Mathematica® is a registered trademark of Mathematica Policy Research